

(19) World Intellectual Property Organization
International Bureau



B5



(43) International Publication Date
18 July 2002 (18.07.2002)

PCT

(10) International Publication Number
WO 02/055701 A2

(51) International Patent Classification⁷: C12N 15/12,
5/10, C12Q 1/68, C07K 14/705, 16/18, G01N 33/50, 33/53

(74) Agents: MANDRAGOURAS, AMY E. et al.; Lahive &
Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(21) International Application Number: PCT/US01/49060

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

(22) International Filing Date:
17 December 2001 (17.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/256,240 15 December 2000 (15.12.2000) US
60/256,588 18 December 2000 (18.12.2000) US
60/258,028 21 December 2000 (21.12.2000) US

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): MIL-
LENNIUM PHARMACEUTICALS, INC. [US/US]; 75
Sidney Street, Cambridge, MA 02139 (US).

Published:

— without international search report and to be republished
upon receipt of that report

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): CURTIS, Rory, A., J.
[GB/US]; 31 Constitution Drive, Southborough, MA 01772
(US).

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: 8099, 46455, 54414, 53736, 67076, 67102, 44181, 67084FL, AND 67084 ALT, HUMAN PROTEINS AND METHODS
OF USE THEREOF

(57) Abstract: The invention provides isolated nucleic acids molecules, designated 8099, 46455, 5441, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules, which encode novel transporter family molecules. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene has been introduced or disrupted. The invention still further provides isolated 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, fusion polypeptides, antigenic peptides and anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

WO 02/055701 A2

8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, AND 67084 ALT,
HUMAN PROTEINS AND METHODS OF USE THEREOF

5 **Related Applications**

This application claims the benefit of prior-filed provisional patent application Serial No. 60/256,240, filed December 15, 2000, entitled "8099 AND 46455, NOVEL HUMAN SUGAR TRANSPORTERS AND USES THEREFOR," prior-filed provisional patent application Serial No. 60/256,588, filed December 18, 2000, entitled "54414 AND 53763,
10 NOVEL HUMAN POTASSIUM CHANNELS AND USES THEREFOR," and prior-filed provisional patent application Serial No. 60/258,028, filed December 21, 2000, entitled "67076, 67102, 44181, 67084FL, and 67084alt, NOVEL HUMAN PHOSPHOLIPID TRANSPORTERS AND USES THEREOF." The entire contents of the above-referenced applications are incorporated herein by this reference.

15

Background of the Invention

Cellular membranes serve to differentiate the contents of a cell from the surrounding environment, and may also serve as effective barriers against the unregulated influx of hazardous or unwanted compounds, and the unregulated efflux of desirable compounds.
20 Membranes are by nature impervious to the unfacilitated diffusion of hydrophilic compounds such as proteins, water molecules, and ions due to their structure: a bilayer of lipid molecules in which the polar head groups face outward (towards the exterior and interior of the cell) and the nonpolar tails face inward (at the center of bilayer, forming a hydrophobic core). Membranes enable a cell to maintain a relatively higher intracellular
25 concentration of desired compounds and a relatively lower intracellular concentration of undesired compounds than are contained within the surrounding environment.

Membranes also present a structural difficulty for cells, in that most desired compounds cannot readily enter the cell, nor can most waste products readily exit the cell through this lipid bilayer. The import and export of such compounds is regulated by
30 proteins which are embedded (singly or in complexes) in the cellular membrane. Two mechanisms exists whereby membrane proteins allow the passage of compounds: non-mediated and mediated transport. Simple diffusion is an example of non-mediated transport, while facilitated diffusion and active transport are examples of mediated transport. Permeases, porters, translocases, translocators, and transporters are proteins that engage in
35 mediated transport (Voet and Voet (1990) Biochemistry, John Wiley and Sons, Inc., New York, N.Y. pp. 484-505).

Sugar transporters are members of the major facilitator superfamily of transporters. These transporters are passive in the sense that they are driven by the substrate concentration

gradient and they exhibit distinct kinetics as well as sugar substrate specificity. Members of this family share several characteristics: (1) they contain twelve transmembrane domains separated by hydrophilic loops; (2) they have intracellular N- and C-termini; and (3) they are thought to function as oscillating pores. The transport mechanism occurs via sugar binding to the exofacial binding site of the transporter, which is thought to trigger a conformational change causing the sugar binding site to re-orient to the endofacial conformation, allowing the release of substrate. These transporters are specific for various sugars and are found in both prokaryotes and eukaryotes. In mammals, sugar transporters transport various monosaccharides across the cell membrane (Walmsley *et al.* (1998) *Trends in Biochem. Sci.* 23:476-481; Barrett *et al.* (1999) *Curr. Op. Cell Biol.* 11:496-502).

At least nine mammalian glucose transporters have been identified, GLUT1 - GLUT9, which are expressed in a tissue-specific manner (*e.g.*, in brain, erythrocyte, kidney, muscle, and adipose tissues) (Shepherd *et al.* (1999) *N. Engl. J. Med.* 341:248-257; Doege *et al.* (2000) *Biochem. J.* 350:771-776). Some GLUT proteins have been shown to be present in low amounts at the plasma membrane during the basal state, at which time large amounts are sequestered in intracellular vesicle stores. Stimulatory molecules specific for each GLUT (such as insulin) regulate the translocation of the GLUT-containing vesicles to the plasma membrane. The vesicles fuse at the membrane and subsequently expose the GLUT protein to the extracellular milieu to allow glucose (and other monosaccharide) transport into the cell (Walmsley *et al.* (1998) *Trends in Biochem. Sci.* 23:476-481; Barrett *et al.* (1999) *Curr. Op. Cell Biol.* 11:496-502). Other GLUT transporters play a role in constitutive sugar transport.

Potassium (K^+) channels are ubiquitous proteins which are involved in the setting of the resting membrane potential as well as in the modulation of the electrical activity of cells. In excitable cells, K^+ channels influence action potential waveforms, firing frequency, and neurotransmitter secretion (Rudy, B. (1988) *Neuroscience*, 25, 729-749; Hille, B. (1992) *Ionic Channels of Excitable Membranes*, 2nd Ed.). In non-excitable cells, they are involved in hormone secretion, cell volume regulation and potentially in cell proliferation and differentiation (Lewis *et al.* (1995) *Annu. Rev. Immunol.*, 13, 623-653). Developments in electrophysiology have allowed the identification and the characterization of an astonishing variety of K^+ channels that differ in their biophysical properties, pharmacology, regulation and tissue distribution (Rudy, B. (1988) *Neuroscience*, 25, 729-749; Hille, B. (1992) *Ionic Channels of Excitable Membranes*, 2nd Ed.). More recently, cloning efforts have shed considerable light on the mechanisms that determine this functional diversity. Furthermore, analyses of structure-function relationships have provided an important set of data concerning the molecular basis of the biophysical properties (selectivity, gating, assembly) and the pharmacological properties of cloned K^+ channels.

Functional diversity of K^+ channels arises mainly from the existence of a great number of genes coding for pore-forming subunits, as well as for other associated regulatory subunits. Two main structural families of pore-forming subunits have been identified. The first one consists of subunits with a conserved hydrophobic core containing six transmembrane domains (TMDs). These K^+ channel α subunits participate in the formation of outward rectifier voltage-gated (Kv) and Ca^{2+} -dependent K^+ channels. The fourth TMD contains repeated positive charges involved in the voltage gating of these channels and hence in their outward rectification (Logothetis *et al.* (1992) *Neuron*, 8, 531-540; Bezanilla *et al.* (1994) *Biophys. J.* 66, 1011-1021).

The second family of pore-forming subunits have only two TMDs. They are essential subunits of inward-rectifying (IRK), G-protein-coupled (GIRK) and ATP-sensitive (KATP) K^+ channels. The inward rectification results from a voltage-dependent block by cytoplasmic Mg^{2+} and polyamines (Matsuda, H. (1991) *Annu. Rev. Physiol.*, 53, 289-298). A conserved domain, called the P domain, is present in all members of both families (Pongs, O. (1993) *J. Membr. Biol.*, 136, 1-8; Heginbotham *et al.* (1994) *Biophys. J.* 66, 1061-1067; Mackinnon, R. (1995) *Neuron*, 14, 889-892; Pascual *et al.*, (1995) *Neuron*, and 14, 1055-1063). This domain is an essential element of the aqueous, K^+ -selective pore. In both groups, the assembly of four subunits is necessary to form a functional K^+ channel (Mackinnon, R. (1991) *Nature*, 350, 232-235; Yang *et al.*, (1995) *Neuron*, 15, 1441-1447).

In both six TMD and two TMD pore-forming subunit families, different subunits coded by different genes can associate to form heterotetramers with new channel properties (Isacoff *et al.*, (1990) *Nature*, 345, 530-534). A selective formation of heteropolymeric channels may allow each cell to develop the best K^+ current repertoire suited to its function. Pore-forming α subunits of Kv channels are classified into different subfamilies according to their sequence similarity (Chandy *et al.* (1993) *Trends Pharmacol. Sci.*, 14: 434). Tetramerization is believed to occur preferentially between members of each subgroup (Covarrubias *et al.* (1991) *Neuron*, 7, 763-773). The domain responsible for this selective association is localized in the N-terminal region and is conserved between members of the same subgroup. This domain is necessary for hetero- but not homo-multimeric assembly within a subfamily and prevents co-assembly between subfamilies. Recently, pore-forming subunits with two TMDs were also shown to co-assemble to form heteropolymers (Duprat *et al.* (1995) *Biochem. Biophys. Res. Commun.*, 212, 657-663. This heteropolymerization seems necessary to give functional GIRKs. IRKs are active as homopolymers but also form heteropolymers.

New structural types of K^+ channels were identified recently in both humans and yeast. These channels have two P domains in their functional subunit instead of only one (Ketchum *et al.* (1995) *Nature*, 376, 690-695; Lesage *et al.* (1996) *J. Biol. Chem.*, 271, 4183-4187; Lesage *et al.* (1996) *EMBO J.*, 15, 1004-1011; Reid *et al.* (1996) *Receptors*

Channels 4, 51-62). The human channel called TWIK-1, has four TMDs. TWIK-1 is expressed widely in human tissues and is particularly abundant in the heart and the brain. TWIK-1 currents are time independent and inwardly rectifying. These properties suggest that TWIK-1 channels are involved in the control of the background K^+ membrane conductance (Lesage *et al.* (1996) *EMBO J.*, 15, 1004-1011).

Potassium channels are potassium ion selective, and can determine membrane excitability (the ability of, for example, a neuron to respond to a stimulus and convert it into an impulse). Potassium channels can also influence the resting potential of membranes, wave forms and frequencies of action potentials, and thresholds of excitation. Potassium channels are typically expressed in electrically excitable cells, *e.g.*, neurons, muscle, endocrine, and egg cells, and may form heteromultimeric structures, *e.g.*, composed of pore-forming and cytoplasmic subunits. Potassium channels may also be found in non-excitable cells, where they may play a role in, *e.g.*, signal transduction. Examples of potassium channels include: (1) the voltage-gated potassium channels, (2) the ligand-gated potassium channels, *e.g.*, neurotransmitter-gated potassium channels, and (3) cyclic-nucleotide-gated potassium channels. Voltage-gated and ligand-gated potassium channels are expressed in the brain, *e.g.*, in brainstem monoaminergic and forebrain cholinergic neurons, where they are involved in the release of neurotransmitters, or in the dendrites of hippocampal and neocortical pyramidal cells, where they are involved in the processes of learning and memory formation. For a detailed description of potassium channels, see Kandel E. R. *et al.*, Principles of Neural Science, second edition, (Elsevier Science Publishing Co., Inc., N.Y. (1985)), the contents of which are incorporated herein by reference.

The E1-E2 ATPase family is a large superfamily of transport enzymes that contains at least 80 members found in diverse organisms such as bacteria, archaea, and eukaryotes (Palmgren, M. G. and Axelsen, K. B., (1998) *Biochim. Biophys. Acta.* 1365:37-45). These enzymes are involved in ATP hydrolysis-dependent transmembrane movement of a variety of inorganic cations (*e.g.*, H^+ , Na^+ , K^+ , Ca^{2+} , Cu^{2+} , Cd^{2+} , and Mg^{2+} ions) across a concentration gradient, whereby the enzyme converts the free energy of ATP hydrolysis into electrochemical ion gradients. E1-E2 ATPases are also known as "P-type" ATPases, referring to the existence of a covalent high-energy phosphoryl-enzyme intermediate in the chemical reaction pathway of these transporters. Until recently, the superfamily contained four major groups: Ca^{2+} transporting ATPases; Na^+/K^+ and gastric H^+/K^+ transporting ATPases; plasma membrane H^+ transporting ATPases of plants, fungi, and lower eukaryotes; and all bacterial P-type ATPases (Kuhlbrandt *et al.* (1998) *Curr. Opin. Struct. Biol.* 8:510-516).

E1-E2 ATPases are phosphorylated at a highly conserved DKTG sequence. Phosphorylation at this site is thought to control the enzyme's substrate affinity. Most E1-E2 ATPases contain ten alpha-helical transmembrane domains, although additional domains

may be present. A majority of known gated-pore translocators contain twelve alpha-helices, including Na⁺/H⁺ antiporters (West (1997) *Biochim. Biophys. Acta* 1331:213-234).

Members of the E1-E2 ATPase superfamily are able to generate electrochemical ion gradients which enable a variety of processes in the cell such as absorption, secretion, transmembrane signaling, nerve impulse transmission, excitation/contraction coupling, and growth and differentiation (Scarborough (1999) *Curr. Op. Cell Biol.* 11:517-522). These molecules are thus critical to normal cell function and well-being of the organism.

Recently, a new class of E1-E2 ATPases was identified, the aminophospholipid transporters or translocators. These transporters transport not cations, but phospholipids (Tang, X. et al. (1996) *Science* 272:1495-1497; Bull, L. N. et al. (1998) *Nat. Genet.* 18:219-224; Mauro, I. et al. (1999) *Biochem. Biophys. Res. Commun.* 257:333-339). These transporters are involved in cellular functions including bile acid secretion and maintenance of the asymmetrical integrity of the plasma membrane.

Given the important biological and physiological roles played by the sugar transporter family of proteins, the potassium channel family of proteins, and the E1-E2 ATPase family of proteins, there exists a need to identify novel potassium channel family members for use in a variety of diagnostic/prognostic, as well as therapeutic applications

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel human sugar transporter family members, referred to herein as "8099 and 46455" nucleic acid and polypeptide molecules. The 8099 and 46455 nucleic acid and polypeptide molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes, e.g., sugar homeostasis. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding 8099 and 46455 polypeptides or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of 8099 and 46455-encoding nucleic acids.

The present invention is also based, at least in part, on the discovery of novel potassium channel family members, referred to herein as "54414 and 53763" nucleic acid and polypeptide molecules. The 54414 and 53763 nucleic acid and protein molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes, e.g., gene expression, intra- or intercellular signaling, and/or membrane excitability or conductance. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding 54414 and 53763 proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of 54414 and 53763-encoding nucleic acids.

The present invention is also based, at least in part, on the discovery of novel human phospholipid transporter family members, referred to herein as "67076, 67102, 44181,

67084FL, or 67084alt" nucleic acid and polypeptide molecules. The 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid and polypeptide molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes, *e.g.*, phospholipid transport (*e.g.*, aminophospholipid transport), absorption, secretion, gene expression, intra- or inter-cellular signaling, and/or cellular proliferation, growth, apoptosis, and/or differentiation. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding 67076, 67102, 44181, 67084FL, or 67084alt polypeptides or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of 67076, 67102, 44181, 67084FL, or 67084alt-encoding nucleic acids.

In one embodiment, the invention features an isolated nucleic acid molecule that includes the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, or SEQ ID NO:27. In another embodiment, the invention features an isolated nucleic acid molecule that encodes a polypeptide including the amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26. In another embodiment, the invention features an isolated nucleic acid molecule that includes the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____, _____, _____, _____, or _____.

In still other embodiments, the invention features isolated nucleic acid molecules including nucleotide sequences that are substantially identical (*e.g.*, 60% identical) to the nucleotide sequence set forth as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, or SEQ ID NO:27. The invention further features isolated nucleic acid molecules including at least 50 contiguous nucleotides of the nucleotide sequence set forth as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, or SEQ ID NO:27. In another embodiment, the invention features isolated nucleic acid molecules which encode a polypeptide including an amino acid sequence that is substantially identical (*e.g.*, 60% identical) to the amino acid sequence set forth as SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26. The present invention also features nucleic acid molecules which encode allelic variants of the

polypeptide having the amino acid sequence set forth as SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26. In addition to isolated nucleic acid molecules encoding full-length polypeptides, the present invention also features nucleic acid molecules which encode fragments, for example, biologically active or antigenic fragments, of the full-length polypeptides of the present invention (e.g., fragments including at least 10 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26). In still other embodiments, the invention features nucleic acid molecules that are complementary to, antisense to, or hybridize under stringent conditions to the isolated nucleic acid molecules described herein.

In another aspect, the invention provides vectors including the isolated nucleic acid molecules described herein (e.g., 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt-encoding nucleic acid molecules). Such vectors can optionally include nucleotide sequences encoding heterologous polypeptides. Also featured are host cells including such vectors (e.g., host cells including vectors suitable for producing 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules and polypeptides).

In another aspect, the invention features isolated 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides and/or biologically active or antigenic fragments thereof. Exemplary embodiments feature a polypeptide including the amino acid sequence set forth as SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, a polypeptide including an amino acid sequence at least 60% identical to the amino acid sequence set forth as SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, a polypeptide encoded by a nucleic acid molecule including a nucleotide sequence at least 60% identical to the nucleotide sequence set forth as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, or SEQ ID NO:27. Also featured are fragments of the full-length polypeptides described herein (e.g., fragments including at least 10 contiguous amino acid residues of the sequence set forth as SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26) as well as allelic variants of the polypeptide having the amino acid sequence set forth as SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26.

The 8099 and 46455 polypeptides and/or biologically active or antigenic fragments thereof, are useful, for example, as reagents or targets in assays applicable to treatment and/or diagnosis of 8099 and 46455 mediated or related disorders. In one embodiment, 8099 and/or 46455 polypeptides or fragments thereof, have an 8099 and/or 46455 activity.
5 In another embodiment, 8099 and/or 46455 polypeptides or fragments thereof, have at least one, preferably two, three, four, five, six, seven, eight, nine, ten, eleven, or twelve transmembrane domains and/or a sugar transporter family domain, and optionally, have an 8099 and/or 46455 activity.

The 54414 and 53763 polypeptides and/or biologically active or antigenic fragments thereof, are useful, for example, as reagents or targets in assays applicable to treatment and/or diagnosis of 54414 and 53763 mediated or related disorders. In one embodiment, a 54414 AND 53763 polypeptide or fragment thereof has a 54414 and 53763 activity. In another embodiment, a 54414 and 53763 polypeptide or fragment thereof has at least one or
10 more of the following domains or motifs: a transmembrane domain, an ion transport protein domain, a K^+ channel tetramerisation domain, a P-loop motif, a pore domain, a potassium channel signature sequence motif, and/or a voltage sensor motif, and optionally, has a 54414 or 53763 activity.
15

The 67076, 67102, 44181, 67084FL, or 67084alt polypeptides and/or biologically active or antigenic fragments thereof, are useful, for example, as reagents or targets in assays applicable to treatment and/or diagnosis of 67076, 67102, 44181, 67084FL, or 67084alt associated or related disorders. In one embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or fragment thereof, has a 67076, 67102, 44181, 67084FL, or 67084alt activity. In another embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or fragment thereof, includes at least one of the following domains, sites, or motifs: a
20 transmembrane domain, an N-terminal large extramembrane domain, a C-terminal large extramembrane domain, an E1-E2 ATPases phosphorylation site, a P-type ATPase sequence 1 motif, a P-type ATPase sequence 2 motif, a P-type ATPase sequence 3 motif, and/or one or more phospholipid transporter specific amino acid residues, and optionally, has a 67076, 67102, 44181, 67084FL, or 67084alt activity.
25

30 In a related aspect, the invention features antibodies (e.g., antibodies which specifically bind to any one of the polypeptides described herein) as well as fusion polypeptides including all or a fragment of a polypeptide described herein.

The present invention further features methods for detecting 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides and/or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules, such
35 methods featuring, for example, a probe, primer or antibody described herein. Also featured are kits, e.g., kits for the detection of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides and/or 8099, 46455, 54414, 53763, 67076, 67102,

44181, 67084FL, or 67084alt nucleic acid molecules. In a related aspect, the invention features methods for identifying compounds which bind to and/or modulate the activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecule described herein. Further featured are methods for modulating a 67076, 67102, 44181, 67084FL, or 67084alt activity.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

10 **Brief Description of the Drawings**

Figures 1A-1B depict the cDNA sequence and predicted amino acid sequence of human 8099. The nucleotide sequence corresponds to nucleic acids 1 to 2725 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 617 of SEQ ID NO:2. The coding region without the 5' and 3' untranslated regions of the human 8099 gene is shown in SEQ ID NO: 3.

Figure 2 depicts a structural, hydrophobicity, and antigenicity analysis of the human 8099 polypeptide (SEQ ID NO:2).

Figure 3A-C depicts the results of a search which was performed against the HMM database in PFAM.

20 *Figure 4* depicts an alignment of the human 8099 amino acid sequence (SEQ ID NO:2) with the amino acid sequence of the *E. coli* galactose-proton symporter GALP using the CLUSTAL W (1.74) alignment program (having GenBank Accession No. P37021, set forth as SEQ ID NO:28).

25 *Figure 5* depicts an alignment of the human 8099 amino acid sequence (SEQ ID NO:2) with the amino acid sequence of the *E. coli* arabinose-proton symporter ARAE using the CLUSTAL W (1.74) alignment program (having GenBank Accession No. P09830, set forth as SEQ ID NO:29).

30 *Figure 6* depicts an alignment of the human 8099 amino acid sequence (SEQ ID NO:2) with the amino acid sequence of *E. coli* GALP and ARAE using the CLUSTAL W (1.74) alignment program (having GenBank Accession Nos. P37021 and P09830, respectively, set forth as SEQ ID NOs:28 and 29, respectively).

Figure 7 depicts an alignment of the human 8099 amino acid sequence (SEQ ID NO:2) with the amino acid sequence of the *H. sapiens* facilitative glucose transporter GLUT8 using the CLUSTAL W (1.74) alignment program (having GenBank Accession No. Y02168, set forth as SEQ ID NO:30).

Figures 8A-B depict the cDNA sequence and predicted amino acid sequence of human 46455. The nucleotide sequence corresponds to nucleic acids 1 to 2230 of SEQ ID NO:4. The amino acid sequence corresponds to amino acids 1 to 528 of SEQ ID NO:5. The coding region without the 5' and 3' untranslated regions of the human 46455 gene is shown in SEQ ID NO: 6.

Figure 9 depicts a structural, hydrophobicity, and antigenicity analysis of the human 46455 polypeptide (SEQ ID NO:5).

Figure 10A-C depicts the results of a search which was performed against the HMM database in PFAM.

Figure 11 depicts an alignment of the human 46455 amino acid sequence (SEQ ID NO:5) with the amino acid sequence of *C. elegans* Z92825 using the CLUSTAL W (1.74) alignment program (having GenBank Accession No. Z92825, set forth as SEQ ID NO:31).

Figure 12A-D depicts the nucleotide sequence of the human 54414 cDNA and the corresponding amino acid sequence. The nucleotide sequence corresponds to nucleic acids 1 to 4632 of SEQ ID NO:7. The amino acid sequence corresponds to amino acids 1 to 1118 of SEQ ID NO:8. The coding region without the 5' or 3' untranslated regions of the human 54414 gene is shown in SEQ ID NO:9.

Figure 13 depicts a structural, hydrophobicity, and antigenicity analysis of the human 54414 polypeptide (SEQ ID NO:8). The locations of the 6 transmembrane domains, as well as the pore domain (P), are indicated.

Figure 14 depicts the results of a search in the HMM database, using the amino acid sequence of human 54414.

Figure 15A-B depicts a Clustal W (1.74) multiple sequence alignment of the human 54414 amino acid sequence (54414.prot; SEQ ID NO:8) and the amino acid sequence of the *Rattus norvegicus* Slack potassium channel subunit (AF089730; SEQ ID NO:32; GenBank Accession No. AAC83350). Amino acid identities are indicated by stars. The six transmembrane domains (TM1, TM2, etc.) are boxed. The pore domain, which contains the potassium channel signature sequence motif, is also boxed.

Figure 16A-C depicts the nucleotide sequence of the human 53763 cDNA and the corresponding amino acid sequence. The nucleotide sequence corresponds to nucleic acids 1 to 2847 of SEQ ID NO:10. The amino acid sequence corresponds to amino acids 1 to 638 of SEQ ID NO:11. The coding region without the 5' or 3' untranslated regions of the human 53763 gene is shown in SEQ ID NO:12.

Figure 17 depicts a structural, hydrophobicity, and antigenicity analysis of the human 53763 polypeptide (SEQ ID NO:11). The locations of the 6 transmembrane domains, as well as the pore domain (P), are indicated.

Figure 18A-B depicts the results of a search in the HMM database, using the amino acid sequence of human 53763.

Figure 19 depicts a Clustal W (1.74) sequence alignment of the human 53763 amino acid sequence (Fbh53763pat; SEQ ID NO:11) and the amino acid sequence of the *Rattus norvegicus* voltage-gated potassium channel protein KV3.2 (KSHIIIA) (ratCIKE; SEQ ID NO:33; GenBank Accession No. P22462). Amino acid identities are indicated by stars.

5 The six transmembrane domains (TM1, TM2, etc.) are boxed. The pore domain, which contains the potassium channel signature sequence motif, is also boxed. Plus signs (+) at every third position of the fourth transmembrane domain (TM4), indicate the positively charged residues of the voltage sensor.

Figures 20A-E depicts the cDNA sequence and predicted amino acid sequence of human 67076. The nucleotide sequence corresponds to nucleic acids 1 to 6582 of SEQ ID NO:13. The amino acid sequence corresponds to amino acids 1 to 1129 of SEQ ID NO:14. The coding region without the 5' and 3' untranslated regions of the human 67076 gene is shown in SEQ ID NO:15.

Figure 21 depicts a structural, hydrophobicity, and antigenicity analysis of the human 67076 polypeptide (SEQ ID NO:14).

Figure 22 depicts the results of a search in the HMM database, using the amino acid sequence of human 67076.

Figures 23 depicts a Clustal W (1.74) alignment of the human 67076 amino acid sequence ("Fbh67076FL"; SEQ ID NO:14) with the amino acid sequence of mouse Potential Phospholipid-Transporting ATPase IH (mouseAT1H) (GenBank Accession No. P98197) (SEQ ID NO:34). The transmembrane domains ("TM1", "TM2", etc.), E1-E2 ATPases phosphorylation site ("phosphorylation site"), and phospholipid transporter specific amino acid residues ("phospholipid transport") are boxed.

Figures 24A-E depicts the cDNA sequence and predicted amino acid sequence of human 67102. The nucleotide sequence corresponds to nucleic acids 1 to 6074 of SEQ ID NO:16. The amino acid sequence corresponds to amino acids 1 to 1426 of SEQ ID NO:17. The coding region without the 5' and 3' untranslated regions of the human 67102 gene is shown in SEQ ID NO:18.

Figure 25 depicts a structural, hydrophobicity, and antigenicity analysis of the human 67102 polypeptide (SEQ ID NO:17).

Figure 26A-B depicts the results of a search in the HMM database, using the amino acid sequence of human 67102.

Figures 27A-B depicts a Clustal W (1.74) alignment of the human 67102 amino acid sequence ("Fbh67102FL"; SEQ ID NO:17) with the amino acid sequence of mouse Potential Phospholipid-Transporting ATPase VA (mouseAT5A) (GenBank Accession No. O54827) (SEQ ID NO:35). The transmembrane domains ("TM1", "TM2", etc.), E1-E2 ATPases phosphorylation site ("phosphorylation site"), and phospholipid transporter specific amino acid residues ("phospholipid transport") are boxed.

Figures 28A-E depicts the cDNA sequence and predicted amino acid sequence of human 44181. The nucleotide sequence corresponds to nucleic acids 1 to 7221 of SEQ ID NO:19. The amino acid sequence corresponds to amino acids 1 to 1177 of SEQ ID NO:20. The coding region without the 5' and 3' untranslated regions of the human 44181 gene is shown in SEQ ID NO:21.

Figure 29 depicts a structural, hydrophobicity, and antigenicity analysis of the human 44181 polypeptide (SEQ ID NO:20).

Figure 30A-B depicts the results of a search in the HMM database, using the amino acid sequence of human 44181.

Figures 31A-B depicts a Clustal W (1.74) multiple sequence alignment of the human 44181 amino acid sequence ("Fbh44181"; SEQ ID NO:20) with the amino acid sequence of mouse Potential Phospholipid-Transporting ATPase IH (mouseAT1H) (GenBank Accession No. P98197) (SEQ ID NO:34) and 67076 ("Fbh67076FL"; SEQ ID NO:14). The transmembrane domains ("TM1", "TM2", etc.), E1-E2 ATPases phosphorylation site ("phosphorylation site"), and phospholipid transporter specific amino acid residues ("phospholipid transport") are boxed.

Figures 32A-D depicts the cDNA sequence and predicted amino acid sequence of human 67084FL. The nucleotide sequence corresponds to nucleic acids 1 to 4198 of SEQ ID NO:22. The amino acid sequence corresponds to amino acids 1 to 1084 of SEQ ID NO:23. The coding region without the 5' and 3' untranslated regions of the human 67084FL gene is shown in SEQ ID NO:24.

Figure 33 depicts a structural, hydrophobicity, and antigenicity analysis of the human 67084FL polypeptide (SEQ ID NO:23).

Figures 34A-B depicts the results of a search in the HMM database, using the amino acid sequence of human 67084FL.

Figures 35A-B depicts a Clustal W (1.74) alignment of the human 67084FL amino acid sequence ("Fbh67084FL"; SEQ ID NO:23) with the amino acid sequence of mouse Potential Phospholipid-Transporting ATPase IIV (mouseAT2B) (GenBank Accession No.:P98195) (SEQ ID NO:36). The transmembrane domains ("TM1", "TM2", etc.), E1-E2 ATPases phosphorylation site ("phosphorylation site"), and phospholipid transporter specific amino acid residues ("phospholipid transport") are boxed.

Figures 36A-D depicts the cDNA sequence and predicted amino acid sequence of human 67084alt. The nucleotide sequence corresponds to nucleic acids 1 to 4231 of SEQ ID NO:25. The amino acid sequence corresponds to amino acids 1 to 1095 of SEQ ID NO:26. The coding region without the 5' and 3' untranslated regions of the human 67084alt gene is shown in SEQ ID NO:27.

Figure 37 depicts a structural, hydrophobicity, and antigenicity analysis of the human 67084alt polypeptide (SEQ ID NO:26).

Figures 38A-B depicts the results of a search in the HMM database, using the amino acid sequence of human 67084.

Figures 39A-B depicts a Clustal W (1.74) alignment of the human 67084alt amino acid sequence ("Fbh67084alt"; SEQ ID NO:26) with the amino acid sequence of mouse Potential Phospholipid-Transporting ATPase IIV (mouseAT2B) (GenBank Accession No.:P98195) (SEQ ID NO:36). The transmembrane domains ("TM1", "TM2", etc.), E1-E2 ATPases phosphorylation site ("phosphorylation site"), and phospholipid transporter specific amino acid residues ("phospholipid transport") are boxed.

10 Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of novel sugar transporter family molecules, referred to herein as "8099 and 46455" nucleic acid and polypeptide molecules. These novel molecules are capable of, for example, modulating a transporter mediated activity (e.g., a sugar transporter mediated activity) in a cell, e.g., a liver cell, fat cell, muscle cell, or blood cell, such as an erythrocyte. These novel molecules are capable of transporting molecules, e.g., hexoses such as D-glucose, D-fructose, D-galactose or mannose across biological membranes and, thus, play a role in or function in a variety of cellular processes, e.g., maintenance of sugar homeostasis. Thus the 8099 and 46455 molecules of the present invention provide novel diagnostic targets and therapeutic agents to control 8099 and 46455-associated disorders, as defined herein.

The present invention is also based, at least in part, on the discovery of novel potassium channel family members, referred to herein as "54414 and 53763" nucleic acid and polypeptide molecules. These novel molecules are capable of, for example, modulating PCH mediated activities in a cell, e.g., a neuronal cell. Thus, the 54414 and 53763 molecules of the present invention provide novel diagnostic targets and therapeutic agents to control 54414 or 53763-associated disorders, as defined herein.

The present invention also is based, at least in part, on the discovery of novel phospholipid transporter family molecules, referred to herein as "67076, 67102, 44181, 67084FL, or 67084alt" nucleic acid and polypeptide molecules. These novel molecules are capable of, for example, transporting phospholipids (e.g., aminophospholipids such as phosphatidylserine and phosphatidylethanolamine, choline phospholipids such as phosphatidylcholine and sphingomyelin, and bile acids) across cellular membranes and, thus, play a role in or function in a variety of cellular processes, e.g., phospholipid transport, absorption, secretion, gene expression, intra- or inter-cellular signaling, and/or cellular proliferation, growth, and/or differentiation. Thus, the 67076, 67102, 44181, 67084FL, and 67084alt molecules of the present invention provide novel diagnostic targets and therapeutic agents to control 67076, 67102, 44181, 67084FL, or 67084alt-associated disorders, as defined herein.

The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin or alternatively, can contain homologues of non-human origin, e.g., rat or mouse proteins. Members of a family can also have common functional characteristics.

10 8099 and 46455 Molecules of the Invention

The family of 8099 and 46455 polypeptides comprise at least one "transmembrane domain" and at least one, preferably two, three, four, five, six, seven, eight, nine, ten, eleven, or twelve transmembrane domains. As used herein, the term "transmembrane domain" includes an amino acid sequence of about 20-45 amino acid residues in length which spans the plasma membrane. More preferably, a transmembrane domain includes about at least 20, 25, 30, 35, 40, or 45 amino acid residues and spans the plasma membrane. Transmembrane domains are rich in hydrophobic residues, and typically have an alpha-helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, e.g., leucines, isoleucines, alanines, valines, phenylalanines, prolines or methionines. Transmembrane domains are described in, for example, Zagotta W.N. *et al*, (1996) *Annual Rev. Neurosci.* 19: 235-263, the contents of which are incorporated herein by reference. A MEMSAT and additional analyses resulted in the identification of twelve transmembrane domains in the amino acid sequence of human 8099 (SEQ ID NO:2) at about residues 32-49, 81-101, 109-130, 138-156, 165-184, 198-217, 279-301, 315-338, 346-364, 463-487, 499-521, and 529-549. A MEMSAT and additional analyses resulted in the identification of twelve transmembrane domains in the amino acid sequence of human 46455 (SEQ ID NO:5) at about residues 58-74, 98-118, 126-145, 165-181, 188-205, 218-238, 273-294, 323-341, 357-377, 386-410, 423-441, and 462-485.

30 Accordingly, 8099 and 46455 polypeptides having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with at least one, preferably at least two, three, four, five, six, seven, eight, nine, ten, eleven, or twelve transmembrane domains of human 8099 and 46455, respectively are within the scope of the invention.

35 Another embodiment of the invention features 8099 molecules which contain an N-terminal unique domain. The term "unique N-terminal domain" as used herein, refers to a protein domain of an 8099 protein family member which includes amino acid residues N-terminal to the sixth transmembrane domain, e.g., the GLUT8-like domain in the amino acid

sequence of the 8099 protein. As used herein, a "unique N-terminal domain" refers to a protein domain which is at least about 150-200 amino acid residues in length, preferably at least about 160-190 amino acid residues in length and shares significantly more sequence homology with about residues 1 to 178 of SEQ ID NO:2 than with about residues 1 to 178 of GLUT8.

Accordingly, 8099 polypeptides having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with at least one unique N-terminal domain of human 8099 (*e.g.*, about amino acids 1-178 of SEQ ID NO:2) are within the scope of the invention.

Yet another aspect of the invention features 8099 proteins having an "extended exofacial loop" between transmembrane domains 9 and 10. Preferably, the first amino acid residue of an extended exofacial loop of 8099 is the first residue C-terminal to the amino acid residues of transmembrane domain 9 and the last residue of the exofacial loop is the first residue N-terminal to the amino acid residues of transmembrane domain 10 of 8099. In a preferred embodiment, an extended exofacial loop is at least about 50-55, 55-60, 60-65, 65-70, 70-75, 75-80, 80-85, 85-90, 90-95, 97 or more amino acid residues in length. For example, in one embodiment, an 8099 protein includes an "extended exofacial loop" of about amino acids 365-462 of SEQ ID NO:2 (97 amino acid residues in length).

Accordingly, 8099 polypeptides having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with at least one extended exofacial loop of human 8099 are within the scope of the invention.

In another embodiment, an 8099 and/or 46455 molecule of the present invention is identified based on the presence of at least one "sugar transporter family domain." As used herein, the term "sugar transporter family domain" includes a protein domain having at least about 300-600 amino acid residues and a sugar transporter mediated activity. Preferably, a sugar transporter family domain includes a polypeptide having an amino acid sequence of about 350-550, 400-550, or more preferably, about 411 or 521 amino acid residues and a sugar transporter mediated activity. To identify the presence of a sugar transporter family domain in an 8099 and/or an 46455 protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein may be searched against a database of known protein domains (*e.g.*, the PFAM HMM database). A PFAM sugar transporter family domain has been assigned the PFAM Accession PF00083. A search was performed against the PFAM HMM database resulting in the identification of a sugar transporter family domain in the amino acid sequence of human 8099 (SEQ ID NO:2) at about residues 43-564 of SEQ ID NO:2. A search was performed against the PFAM HMM database resulting in the identification of a sugar transporter family domain in the amino acid sequence of human 46455 (SEQ ID NO:5) at about residues 58-487 of SEQ ID NO:5.

Preferably a "sugar transporter family domain" has a "sugar transporter mediated activity" as described herein. For example, a sugar transporter family domain may have the ability to bind a monosaccharide (e.g., D-glucose, D-fructose, D-galactose and/or mannose); the ability to transport a monosaccharide (e.g., D-glucose, D-fructose, D-galactose and/or mannose) in a constitutive manner or in response to stimuli (e.g., insulin) across a cell membrane (e.g., a liver cell membrane, fat cell membrane, muscle cell membrane, and/or blood cell membrane, such as an erythrocyte membrane); the ability to function as a neuronal transporter; the ability to mediate trans-epithelial movement; and/or the ability to modulate sugar homeostasis in a cell. Accordingly, identifying the presence of a "sugar transporter family domain" can include isolating a fragment of an 8099 and/or an 46455 molecule (e.g., an 8099 and/or an 46455 polypeptide) and assaying for the ability of the fragment to exhibit one of the aforementioned sugar transporter mediated activities.

A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28:405-420 and a detailed description of HMMs can be found, for example, in Gribkov *et al.* (1990) *Meth. Enzymol.* 183:146-159; Gribkov *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.* (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.* (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference.

In a preferred embodiment, the 8099 and/or 46455 molecules of the invention include at least one, preferably two, even more preferably at least three, four, five, six, seven, eight, nine, ten, eleven, or twelve transmembrane domain(s) and/or at least one sugar transporter family domain. In another preferred embodiment, the 8099 molecules of the invention include at least one, preferably two, even more preferably at least three, four, five, six, seven, eight, nine, ten, eleven, or twelve transmembrane domain(s), at least one sugar transporter family domain, at least one unique N-terminal domain, and/or at least one extended exofacial loop.

Isolated polypeptides of the present invention, preferably 8099 or 46455 polypeptides, have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:2 or 5 or are encoded by a nucleotide sequence sufficiently identical to SEQ ID NO:1, 3, 4 or 6. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homology or identity across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently identical. Furthermore, amino acid or nucleotide

sequences which share at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homology or identity and share a common functional activity are defined herein as sufficiently identical.

In a preferred embodiment, an 8099 and/or 46455 polypeptide includes at least one or more of the following domains: a transmembrane domain and/or a sugar transporter family domain, and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous or identical to the amino acid sequence of SEQ ID NO:2 or 5, or the amino acid sequences encoded by the DNA inserts of the plasmids deposited with ATCC as Accession Numbers _____ and/or _____. In yet another preferred embodiment, an 8099 and/or an 46455 polypeptide includes at least one or more of the following domains: a transmembrane domain and/or a sugar transporter family domain, and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4 or SEQ ID NO:6. In another preferred embodiment, an 8099 and/or an 46455 polypeptide includes at least one or more of the following domains: a transmembrane domain and/or a sugar transporter family domain, and has an 8099 and/or an 46455 activity.

As used interchangeably herein, an "8099 activity", "46455 activity", "biological activity of 8099", "biological activity of 46455", "functional activity of 8099" or "functional activity of 46455" refers to an activity exerted by an 8099 and/or 46455 polypeptide or nucleic acid molecule on an 8099 and/or 46455 responsive cell or tissue, or on an 8099 and/or 46455 polypeptide substrate, as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, an 8099 and/or 46455 activity is a direct activity, such as an association with an 8099- and/or 46455-target molecule. As used herein, a "substrate," "target molecule," or "binding partner" is a molecule with which an 8099 and/or 46455 polypeptide binds or interacts in nature, such that 8099- and/or 46455-mediated function is achieved. An 8099 and/or 46455 target molecule can be a non- 8099 and/or a non-46455 molecule or an 8099 and/or 46455 polypeptide or polypeptide of the present invention. In an exemplary embodiment, an 8099 and/or 46455 target molecule is an 8099 and/or 46455 ligand, e.g., a sugar transporter ligand such D-glucose, D-fructose, D-galactose, and/or mannose. Alternatively, an 8099 and/or 46455 activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the 8099 and/or 46455 polypeptide with an 8099 and/or 46455 ligand. The biological activities of 8099 and/or 46455 are described herein. For example, the 8099 and/or 46455 polypeptides of the present invention can have one or more of the following activities: (1) bind a monosaccharide, e.g., D-glucose, D-fructose, D-galactose, and/or mannose, (2) transport monosaccharides across a cell membrane, (3) influence insulin and/or glucagon secretion, (4) maintain sugar homeostasis in a cell, (5) function as a neuronal transporter, and (6) mediate trans-epithelial movement in

a cell. Moreover, in a preferred embodiment, 8099 and/or 46455 molecules of the present invention, 8099 and/or 46455 antibodies, 8099 and/or 46455 modulators are useful in at least one of the following: (1) modulation of insulin sensitivity; (2) modulation of blood sugar levels; (3) treatment of blood sugar level disorders (e.g., diabetes); and/or (4) modulation of insulin resistance.

The nucleotide sequence of the isolated human 8099 and 46455 cDNAs and the predicted amino acid sequences of the human 8099 and 46455 polypeptides are shown in Figures 1 and 8 and in SEQ ID NOs:1 and 2, and SEQ ID NOs:4 and 5, respectively. Plasmids containing the nucleotide sequences encoding human 8099 or 46455 were deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Numbers _____ or _____. These deposits will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. These deposits were made merely as a convenience for those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112.

The human 8099 gene, which is approximately 2725 nucleotides in length, encodes a polypeptide which is approximately 617 amino acid residues in length. The human 46455 gene, which is approximately 2230 nucleotides in length, encodes a polypeptide which is approximately 528 amino acid residues in length.

54414 and 53763 Molecules of the Invention

The family of 54414 and 53763 proteins of the present invention comprises at least one transmembrane domain, preferably at least 2 or 3 transmembrane domains, more preferably 4 or 5 transmembrane domains, and most preferably, 6 transmembrane domains. Amino acid residues 64-83, 104-127, 135-153, 161-173, 199-217, and 257-274 of the human 54414 protein (SEQ ID NO:8) are predicted to comprise transmembrane domains. Amino acid residues 230-248, 287-303, 314-335, 346-368, 382-402, and 451-473 of the human 53763 protein (SEQ ID NO:11) are predicted to comprise transmembrane domains.

In another embodiment, members of the 54414 and 53763 family of proteins include at least one "ion transport protein domain" in the protein or corresponding nucleic acid molecule. As used herein, the term "ion transport protein domain" includes a protein domain having at least about 150-310 amino acid residues and a bit score of at least 200 when compared against an ion transport protein domain Hidden Markov Model (HMM), e.g., PFAM Accession Number PF00520. Preferably, an ion transport protein domain includes a protein domain having an amino acid sequence of about 170-290, 190-270, 210-250, or more preferably about 173 or 191 amino acid residues. To identify the presence of an ion transport protein domain in a 54414 or 53763 protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein is

searched against a database of known protein motifs and/or domains (*e.g.*, the HMM database). The ion transport protein domain (HMM) has been assigned the PFAM Accession number PF00520. A search was performed against the HMM database resulting in the identification of an ion transport protein domain in the amino acid sequence of human 54414 at about residues 104-277 of SEQ ID NO:8 and in the amino acid sequence of human 53763 about residues 281-472 of SEQ ID NO:11.

Preferably an ion transport protein domain is at least about 150-310 amino acid residues and has an "ion transport protein domain activity", for example, the ability to interact with a 54414 or 53763 substrate or target molecule (*e.g.*, a potassium ion) and/or to regulate 54414 or 53763 activity. Accordingly, identifying the presence of an "ion transport protein domain" can include isolating a fragment of a 54414 or 53763 molecule (*e.g.*, a 54414 or 53763 polypeptide) and assaying for the ability of the fragment to exhibit one of the aforementioned ion transport protein domain activities.

In another embodiment, members of the 54414 and 53763 family of proteins include at least one "K⁺ channel tetramerisation domain" in the protein or corresponding nucleic acid molecule. As used herein, the term "K⁺ channel tetramerisation domain" includes a protein domain having at least about 70-230 amino acid residues and a bit score of at least 80 when compared against a K⁺ channel tetramerisation domain Hidden Markov Model (HMM), *e.g.*, PFAM Accession Number PF02214. Preferably, a K⁺ channel tetramerisation domain includes a protein domain having an amino acid sequence of about 90-210, 110-190, 130-170, or more preferably about 149 amino acid residues, and a bit score of at least 100, 120, 140, or more preferably, 156.7. To identify the presence of a K⁺ channel tetramerisation domain in a 54414 or 53763 protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein is searched against a database of known protein motifs and/or domains (*e.g.*, the HMM database). The K⁺ channel tetramerisation domain (HMM) has been assigned the PFAM Accession number PF02214. A search was performed against the HMM database resulting in the identification of a K⁺ channel tetramerisation domain in the amino acid sequence of human 53763 at about residues 8-156 of SEQ ID NO:11.

Preferably a K⁺ channel tetramerisation domain is at least about 70-230 amino acid residues and has an "K⁺ channel tetramerisation domain activity", for example, the ability to interact with one or more potassium channel subunits (*e.g.*, 54414 or 53763 molecules, or non-54414 or 53763 potassium channel subunits), the ability to regulate assembly of a 54414 or 53763 molecule into a potassium channel tetramer, and/or to regulate 54414 or 53763 activity. Accordingly, identifying the presence of an "K⁺ channel tetramerisation domain" can include isolating a fragment of a 54414 or 53763 molecule (*e.g.*, a 54414 or 53763 polypeptide) and assaying for the ability of the fragment to exhibit one of the aforementioned K⁺ channel tetramerisation domain activities.

In another embodiment, a 54414 or 53763 protein of the present invention is identified based on the presence of an "ATP/GTP-binding site motif A (P-loop) motif", referred to alternatively herein as a "P-loop motif", in the protein or corresponding nucleic acid molecule. Preferably, a P-loop motif includes a protein motif which is about 4-15, 5-13, 6-11, 7-9, or preferably about 8 amino acid residues. The P-loop motif functions in binding ATP and/or GTP via interaction with the phosphate groups of the nucleotide and has been assigned Prosite™ Accession Number PS00017. To identify the presence of a P-loop motif in a 54414 or 53763 protein, and to make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein may be searched against a database of known protein domains or motifs (e.g., the Prosite™ database) using the default parameters (available at the ProSite website). A search was performed against the ProSite database resulting in the identification of a P-loop motif in the amino acid sequence of human 54414 (SEQ ID NO:8) at about residues 1007-1014.

In another embodiment, a 54414 or 53763 protein of the present invention is identified based on the presence of a "pore domain", alternatively referred to herein as a "P-region domain", in the protein or corresponding nucleic acid molecule. As used interchangeably herein, the terms "pore domain" and "P-region domain" include a protein domain having about 10-30, 12-28, 13-25, 14-24, 15-23, or preferably about 16-22 amino acid residues, which is involved in lining the potassium channel pore. A pore domain is typically found between transmembrane domains of potassium channels and is believed to be a major determinant of ion selectivity in potassium channels. Preferably, a pore domain includes a potassium channel signature motif, as defined herein. Pore domains are described in, for example, Warmke et al. (1991) *Science* 252:1560-1562; Zagotta W.N. et al. (1996) *Annu. Rev. Neurosci.* 19:235-63; Pongs, O. (1993) *J. Membr. Biol.* 136:1-8; Heginbotham et al. (1994) *Biophys. J.* 66:1061-1067; Mackinnon, R. (1995) *Neuron* 14:889-892; and Pascual et al. (1995) *Neuron* 14:1055-1063), the contents of which are incorporated herein by reference. A pore domain was identified in the amino acid sequence of human 54414 at about residues 229-250 of SEQ ID NO:8. A pore domain was identified in the amino acid sequence of human 53763 at about residues 426-441 of SEQ ID NO:11.

In a further embodiment, a 54414 or 53763 protein of the present invention is identified based on the presence of a "potassium channel signature sequence motif" in the protein or corresponding nucleic acid molecule. As used herein, the term "potassium channel signature sequence motif" includes a protein motif which is diagnostic for potassium channels. Preferably, a potassium channel signature sequence motif has the consensus sequence T-X-X-T-X-G-hydrophobic-G (see Joiner, W. J. et al. (1998) *Nat. Neurosci.* 1:462-469 and references cited therein), wherein "X" indicates any amino acid residue, and "hydrophobic" indicates any hydrophobic amino acid residue. Preferably, a potassium channel signature sequence motif is included within a pore domain and includes

at least 1, 2, 3, 4, 5, 6, 7, or more preferably, 8 amino acid residues that match the consensus sequence for a potassium channel signature sequence motif. A potassium channel signature sequence motif was identified in the amino acid sequence of human 54414 at about residues 239-246 of SEQ ID NO:8. A potassium channel signature sequence motif was identified in the amino acid sequence of human 53763 at about residues 436-441 of SEQ ID NO:11.

In still another embodiment, a 54414 or 53763 protein of the present invention is identified based on the presence of a "voltage sensor motif", alternatively referred to simply as a "voltage sensor", in the protein or the corresponding nucleic acid molecule. As used interchangeably herein, the terms "voltage sensor motif" and "voltage sensor" include a protein motif having about 10-30, 11-26, 12-24, 13-22, 14-20, 15-18, or more preferably 16 amino acid residues, which is involved in sensing voltage differences between the two sides of the plasma membrane of a cell. Preferably, a voltage sensor motif includes at least 1, 2, 3, 4, 5, or more preferably, 6 positively charged amino acid residues, which are preferably spaced apart by at least 1, or preferably 2, non-positively charged amino acid residues.

Preferably, a voltage sensor motif is included within and/or overlaps with a transmembrane domain, more preferably the fourth transmembrane, of the 54414 or 53763 protein in which it is found. A voltage sensor motif was identified in the amino acid sequence of human 53763 at about residues 348-363 of SEQ ID NO:8. The positively charged amino acid residues of the human 53763 voltage sensor were identified at about residues 348, 351, 354, 357, 360, and 363 of SEQ ID NO:8. No voltage sensor was identified in human 54414.

Isolated proteins of the present invention, preferably 54414 or 53763 proteins, have an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:8 or SEQ ID NO:11, or are encoded by a nucleotide sequence sufficiently homologous to SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, or SEQ ID NO:12. Amino acid or nucleotide sequences which share common structural domains having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homology or identity across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homology or identity and share a common functional activity are defined herein as sufficiently homologous.

In a preferred embodiment, a 54414 or 53763 protein includes at least one or more of the following domains or motifs: a transmembrane domain, an ion transport protein domain, a K⁺ channel tetramerisation domain, a P-loop motif, a pore domain, a potassium channel signature sequence motif, and/or a voltage sensor motif. and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous or identical to the amino acid sequence of SEQ

ID NO:8 or 11, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____ or _____. In yet another preferred embodiment, a 54414 or 53763 protein includes at least one or more of the following domains or motifs: a transmembrane domain, an ion transport protein domain, a K⁺ channel tetramerisation domain, a P-loop motif, a pore domain, a potassium channel signature sequence motif, and/or a voltage sensor motif, and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:7, 9, 10, or 12. In another preferred embodiment, a 54414 or 53763 protein includes at least one or more of the following domains or motifs: a transmembrane domain, an ion transport protein domain, a K⁺ channel tetramerisation domain, a P-loop motif, a pore domain, a potassium channel signature sequence motif, and/or a voltage sensor motif, and has a 54414 or 53763 activity.

As used interchangeably herein, a "54414 or 53763 activity", "biological activity of 54414 or 53763" or "functional activity of 54414 or 53763", includes an activity exerted or mediated by a 54414 or 53763 protein, polypeptide or nucleic acid molecule when expressed in a cell or on a membrane, as determined *in vivo* or *in vitro*, according to standard techniques. In one embodiment, a 54414 or 53763 activity is a direct activity, such as transport of a 54414 or 53763 substrate (*e.g.*, a potassium ion). In another embodiment, a 54414 or 53763 activity is an indirect activity mediated, for example, by interaction of a 54414 or 53763 molecule with a 54414 or 53763 target molecule or binding partner. As used herein, a "target molecule" or "binding partner" is a molecule with which a 54414 or 53763 protein binds or interacts in nature, such that function of the target molecule or binding partner is modulated. In an exemplary embodiment, a 54414 or 53763 target molecule or binding partner is a 54414 or 53763 polypeptide or a non-54414 or 53763 potassium channel subunit.

In a preferred embodiment, a 54414 or 53763 activity is at least one of the following activities: (i) interaction with a 54414 or 53763 substrate (*e.g.*, a potassium ion or a cyclic nucleotide); (ii) conductance or transport of a 54414 or 53763 substrate across a cellular membrane; (iii) interaction with a second protein (*e.g.*, a second 54414 or 53763 subunit or a non-54414 or 53763 potassium channel subunit); (iv) modulation (*e.g.*, maintenance and/or rectification) of membrane potentials; (v) regulation of target molecule availability or activity; (vi) modulation of intra- or intercellular signaling and/or gene transcription (*e.g.*, either directly or indirectly); (viii) generation of outwardly rectifying currents; (viii) modulation of membrane excitability; (ix) modulation of the release of neurotransmitters; (x) regulation of contractility (*e.g.*, of smooth muscle cells), secretion, and/or synaptic transmission; and/or (xi) modulation of processes which underlie learning and memory.

Preferred activities of 54414 further include at least one of the following activities:
(i) interaction with maxi-K potassium channels (*i.e.*, large conductance channels, in particular *Slo*); (ii) modulation of maxi-K potassium channel activity (*e.g.*, *Slo*-mediated activities); (iii) generation of intermediate conductance channels; and/or (iv) regulation of contractility (*e.g.*, of smooth muscle cells), secretion, and/or synaptic transmission, in particular, via modulation of *Slo*.

Preferred activities of 53763 further include at least one of the following activities:
(i) interaction with Shaker (Sh) potassium channels and/or channel subunits; (ii) modulation of Shaker (Sh) potassium channel activity (*e.g.*, termination of prolonged membrane depolarization; (iii) modulation of high voltage activating channel activity and/or inactivating channel activity, and the like.

The nucleotide sequence of the isolated human 54414 cDNA and the predicted amino acid sequence encoded by the 54414 cDNA are shown in Figures 12A-C and in SEQ ID NOs:7 and 8, respectively. A plasmid containing the human 54414 cDNA was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit were made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human 54414 gene, which is approximately 4632 nucleotides in length, encodes a protein having a molecular weight of approximately 123 kD and which is approximately 1118 amino acid residues in length.

The nucleotide sequence of the isolated human 53763 cDNA and the predicted amino acid sequence encoded by the 53763 cDNA are shown in Figures 16A-C and in SEQ ID NOs:10 and 11, respectively. A plasmid containing the human 53763 cDNA was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit were made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human 53763 gene, which is approximately 2847 nucleotides in length, encodes a protein having a molecular weight of approximately 70.2 kD and which is approximately 638 amino acid residues in length.

67076, 67102, 44181, 67084FL, and 67084alt Molecules of the Invention

The 67076, 67102, 44181, 67084FL, and 67084alt polypeptides comprise at least one "transmembrane domain" and preferably eight, nine, or ten transmembrane domains. A MEMSAT analysis and a structural, hydrophobicity, and antigenicity analysis also resulted in the identification of ten transmembrane domains in the amino acid sequence of human 67076 (SEQ ID NO:14) at about residues 57-77, 84-105, 292-313, 345-365, 863-883, 905-926, 956-977, 989-1009, 1021-1041, and 1060-1087. A MEMSAT analysis and a structural, hydrophobicity, and antigenicity analysis resulted in the identification of ten transmembrane domains in the amino acid sequence of human 67102 (SEQ ID NO:17) at about residues 98-115, 122-140, 322-344, 366-390, 582-601, 752-770, 1145-1166, 1225-1246, 1253-1276, and 1298-1317. A MEMSAT analysis and a structural, hydrophobicity, and antigenicity analysis resulted in the identification of ten transmembrane domains in the amino acid sequence of human 44181 (SEQ ID NO:20) at about residues 56-72, 87-103, 290-311, 343-363, 878-898, 911-931, 961-982, 995-1015, 1027-1047, and 1062-1086. A MEMSAT analysis and a structural, hydrophobicity, and antigenicity analysis resulted in the identification of ten transmembrane domains in the amino acid sequence of human 67084FL (SEQ ID NO:23) at about residues 104-120, 124-144, 331-350, 357-374, 887-903, 912-931, 961-983, 990-1008, 1015-1035, and 1043-1067. A MEMSAT analysis and a structural, hydrophobicity, and antigenicity analysis resulted in the identification of ten transmembrane domains in the amino acid sequence of human 67084alt (SEQ ID NO:26) at about residues 104-120, 124-144, 331-350, 357-379, 887-903, 912-931, 961-983, 990-1008, 1015-1035, and 1054-1078.

The family of 67076, 67102, 44181, 67084FL, or 67084alt proteins of the present invention also comprises at least one "extramembrane domain" in the protein or corresponding nucleic acid molecule. As used herein, an "extramembrane domain" includes a domain having greater than 20 amino acid residues that is found between transmembrane domains, preferably on the cytoplasmic side of the plasma membrane, and does not span or traverse the plasma membrane. An extramembrane domain preferably includes at least one, two, three, four or more motifs or consensus sequences characteristic of P-type ATPases, *i.e.*, includes one, two, three, four, or more "P-type ATPase consensus sequences or motifs". As used herein, the phrase "P-type ATPase consensus sequences or motifs" includes any consensus sequence or motif known in the art to be characteristic of P-type ATPases, including, but not limited to, the P-type ATPase sequence 1 motif (as defined herein), the P-type ATPase sequence 2 motif (as defined herein), the P-type ATPase sequence 3 motif (as defined herein), and the E1-E2 ATPases phosphorylation site (as defined herein).

In one embodiment, the family of 67076, 67102, 44181, 67084FL, or 67084alt proteins of the present invention comprises at least one "N-terminal" large extramembrane domain in the protein or corresponding nucleic acid molecule. As used herein, an "N-

terminal" large extramembrane domain is found in the N-terminal 1/3rd of the protein, preferably between the second and third transmembrane domains of a 67076, 67102, 44181, 67084FL, or 67084alt protein and includes about 60-300, 80-280, 100-260, 120-240, 140-220, 160-200, or preferably, 180, 185, or 186 amino acid residues. In a preferred embodiment, an N-terminal large extramembrane domain includes at least one P-type ATPase sequence 1 motif (as described herein). An N-terminal large extramembrane domain was identified in the amino acid sequence of human 67076 at about residues 106-291 of SEQ ID NO:14. An N-terminal large extramembrane domain was identified in the amino acid sequence of human 67102 at about residues 141-321 of SEQ ID NO:17. An N-terminal large extramembrane domain was identified in the amino acid sequence of human 44181 at about residues 104-289 of SEQ ID NO:20. An N-terminal large extramembrane domain was identified in the amino acid sequence of human 67084FL at about residues 145-330 of SEQ ID NO:23. An N-terminal large extramembrane domain was identified in the amino acid sequence of human 67087alt at about residues 145-330 of SEQ ID NO:26.

The family of 67076, 67102, 44181, 67084FL, or 67084alt proteins of the present invention also comprises at least one "C-terminal" large extramembrane domain in the protein or corresponding nucleic acid molecule. As used herein, a "C-terminal" large extramembrane domain is found in the C-terminal 2/3rds of the protein, preferably between the fourth and fifth transmembrane domains of a 67076, 67102, 44181, 67084FL, or 67084alt protein and includes about 150-1000, 300-900, 370-850, 400-820, 430-790, 460-760, 430-730, 460-700, 430-670, 460-640, 430-610, 490-580, 510-550, or preferably, 190, 506, or 523 amino acid residues. In a preferred embodiment, a C-terminal large extramembrane domain includes at least one or more of the following motifs: a P-type ATPase sequence 2 motif (as described herein), a P-type ATPase sequence 3 motif (as defined herein), and/or an E1-E2 ATPases phosphorylation site (as defined herein). A C-terminal large extramembrane domain was identified in the amino acid sequence of human 67076 at about residues 366-862 of SEQ ID NO:14. A C-terminal large extramembrane domain was identified in the amino acid sequence of human 67102 at about residues 391-581 of SEQ ID NO:17. A C-terminal large extramembrane domain was identified in the amino acid sequence of human 44181 at about residues 364-877 of SEQ ID NO:20. A C-terminal large extramembrane domain was identified in the amino acid sequence of human 67084FL at about residues 380-886 of SEQ ID NO:23. A C-terminal large extramembrane domain was identified in the amino acid sequence of human 67084alt at about residues 380-886 of SEQ ID NO:26.

In another embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt protein or 67076, 67102, 44181, 67084FL, or 67084alt extramembrane domain is characterized by at least one "P-type ATPase sequence 1 motif" in the protein or corresponding nucleic acid sequence. As used herein, a "P-type ATPase sequence 1 motif" is a conserved sequence

motif diagnostic for P-type ATPases (Tang, X. *et al.* (1996) *Science* 272:1495-1497; Fagan, M. J. and Saier, M. H. (1994) *J. Mol. Evol.* 38:57). Amino acid residues of the P-type ATPase sequence 1 motif are involved in the coupling of ATP hydrolysis with transport (e.g., transport of phospholipids). The consensus sequence for a P-type ATPase sequence 1 motif is [DNS]-[QENR]-[SA]-[LIVSAN]-[LIV]-[TSN]-G-E-[SN] (SEQ ID NO:37). The use of amino acids in brackets indicates that the amino acid at the indicated position may be any one of the amino acids within the brackets; e.g., [SA] indicates any of one of either S (serine) or A (alanine). In a preferred embodiment, a P-type ATPase sequence 1 motif is contained within an N-terminal large extramembrane domain. In another preferred embodiment, a P-type ATPase sequence 1 motif in the 67076, 67102, 44181, 67084FL, or 67084alt proteins of the present invention has at least 1, 2, 3, or preferably 4 amino acid residues which match the consensus sequence for a P-type ATPase sequence 1 motif. A P-type ATPase sequence 1 motif was identified in the amino acid sequence of human 67076 at about residues 173-181 of SEQ ID NO:14. A P-type ATPase sequence 1 motif was identified in the amino acid sequence of human 67102 at about residues 208-216 of SEQ ID NO:17. A P-type ATPase sequence 1 motif was identified in the amino acid sequence of human 44181 at about residues 173-181 of SEQ ID NO:20. A P-type ATPase sequence 1 motif was identified in the amino acid sequence of human 67084FL at about residues 213-221 of SEQ ID NO:23. A P-type ATPase sequence 1 motif was identified in the amino acid sequence of human 67084alt at about residues 213-221 of SEQ ID NO:26.

In another embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt protein or 67076, 67102, 44181, 67084FL, or 67084alt extramembrane domain is characterized by at least one "P-type ATPase sequence 2 motif" in the protein or corresponding nucleic acid sequence. As used herein, a "P-type ATPase sequence 2 motif" is a conserved sequence motif diagnostic for P-type ATPases (Tang, X. *et al.* (1996) *Science* 272:1495-1497; Fagan, M. J. and Saier, M. H. (1994) *J. Mol. Evol.* 38:57). Preferably, a P-type ATPase sequence 2 motif overlaps with and/or includes an E1-E2 ATPases phosphorylation site (as defined herein). The consensus sequence for a P-type ATPase sequence 2 motif is [LIV]-[CAML]-[STFL]-D-K-T-G-T-[LI]-T (SEQ ID NO:38). The use of amino acids in brackets indicates that the amino acid at the indicated position may be any one of the amino acids within the brackets; e.g., [LI] indicates any of one of either L (leucine) or I (isoleucine). In a preferred embodiment, a P-type ATPase sequence 2 motif is contained within a C-terminal large extramembrane domain. In another preferred embodiment, a P-type ATPase sequence 2 motif in the 67076, 67102, 44181, 67084FL, or 67084alt proteins of the present invention has at least 1, 2, 3, 4, 5, 6, 7, 8, or more preferably 9 amino acid residues which match the consensus sequence for a P-type ATPase sequence 2 motif. A P-type ATPase sequence 2 motif was identified in the amino acid sequence of human 67076 at about residues 406-415 of SEQ ID NO:14. A P-type ATPase sequence 2 motif was identified in the amino acid

sequence of human 67102 at about residues 435-444 of SEQ ID NO:17. A P-type ATPase sequence 2 motif was identified in the amino acid sequence of human 44181 at about residues 404-413 of SEQ ID NO:20. A P-type ATPase sequence 2 motif was identified in the amino acid sequence of human 67084FL at about residues 413-422 of SEQ ID NO:23.

5 A P-type ATPase sequence 2 motif was identified in the amino acid sequence of human 67084alt at about residues 413-422 of SEQ ID NO:26.

In yet another embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt protein or 67076, 67102, 44181, 67084FL, or 67084alt extramembrane domain is characterized by at least one "P-type ATPase sequence 3 motif" in the protein or corresponding nucleic acid

10 sequence. As used herein, a "P-type ATPase sequence 3 motif" is a conserved sequence motif diagnostic for P-type ATPases (Tang, X. et al. (1996) *Science* 272:1495-1497; Fagan, M. J. and Saier, M. H. (1994) *J. Mol. Evol.* 38:57). Amino acid residues of the P-type ATPase sequence 3 motif are involved in ATP binding. The consensus sequence for a P-type ATPase sequence 3 motif is [TIV]-G-D-G-X-N-D-[ASG]-P-[ASV]-L (SEQ ID

15 NO:39). X indicates that the amino acid at the indicated position may be any amino acid (i.e., is not conserved). The use of amino acids in brackets indicates that the amino acid at the indicated position may be any one of the amino acids within the brackets, e.g., [TIV] indicates any of one of either T (threonine), I (isoleucine), or V (valine). In a preferred embodiment, a P-type ATPase sequence 3 motif is contained within a C-terminal large

20 extramembrane domain. In another preferred embodiment, a P-type ATPase sequence 3 motif in the 67076, 67102, 44181, 67084FL, or 67084alt proteins of the present invention has at least 1, 2, 3, 4, 5, 6, or more preferably 7 amino acid residues (including the amino acid at the position indicated by "X") which match the consensus sequence for a P-type ATPase sequence 3 motif. A P-type ATPase sequence 3 motif was identified in the amino acid

25 sequence of human 67076 at about residues 813-824 of SEQ ID NO:14. A P-type ATPase sequence 3 motif was identified in the amino acid sequence of human 67102 at about residues 1054-1064 of SEQ ID NO:17. A P-type ATPase sequence 3 motif was identified in the amino acid sequence of human 44181 at about residues 819-829 of SEQ ID NO:20. A P-type ATPase sequence 3 motif was identified in the amino acid sequence of human

30 67084FL at about residues 820-830 of SEQ ID NO:23. A P-type ATPase sequence 3 motif was identified in the amino acid sequence of human 67084alt at about residues 820-830 of SEQ ID NO:26.

In another embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt protein of the present invention is identified based on the presence of an "E1-E2 ATPases phosphorylation

35 site" (alternatively referred to simply as a "phosphorylation site") in the protein or corresponding nucleic acid molecule. An E1-E2 ATPases phosphorylation site functions in accepting a phosphate moiety and has the amino acid sequence DKTGT (amino acid residues 4-8 of SEQ ID NO:38), and can be included within the E1-E2 ATPase

phosphorylation site consensus sequence: D-K-T-G-T-[LIVM]-[TI] (SEQ ID NO:41), wherein D is phosphorylated. The use of amino acids in brackets indicates that the amino acid at the indicated position may be any one of the amino acids within the brackets, *e.g.*, [TI] indicates any of one of either T (threonine) or I (isoleucine). The E1-E2 ATPases phosphorylation site consensus sequence has been assigned ProSite Accession Number PS00154. To identify the presence of an E1-E2 ATPases phosphorylation site consensus sequence in a 67076, 67102, 44181, 67084FL, or 67084alt protein, and to make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein may be searched against a database of known protein motifs (*e.g.*, the ProSite database) using the default parameters (available at the Prosite website). A search was performed against the ProSite database resulting in the identification of an E1-E2 ATPases phosphorylation site consensus sequence in the amino acid sequence of human 67076 (SEQ ID NO:14) at about residues 409-415. A search was performed against the ProSite database resulting in the identification of an E1-E2 ATPases phosphorylation site consensus sequence in the amino acid sequence of human 67102 (SEQ ID NO:17) at about residues 438-444. A search was performed against the ProSite database resulting in the identification of an E1-E2 ATPases phosphorylation site consensus sequence in the amino acid sequence of human 44181 (SEQ ID NO:20) at about residues 407-413. A search was performed against the ProSite database resulting in the identification of an E1-E2 ATPases phosphorylation site consensus sequence in the amino acid sequence of human 67084FL (SEQ ID NO:23) at about residues 416-422. A search was performed against the ProSite database resulting in the identification of an E1-E2 ATPases phosphorylation site consensus sequence in the amino acid sequence of human 67084alt (SEQ ID NO:26) at about residues 416-422.

Preferably an E1-E2 ATPases phosphorylation site has a "phosphorylation site activity," for example, the ability to be phosphorylated; to be dephosphorylated; to regulate the E1-E2 conformational change of the phospholipid transporter in which it is contained; to regulate transport of phospholipids (*e.g.*, aminophospholipids such as phosphatidylserine and phosphatidylethanolamine, choline phospholipids such as phosphatidylcholine and sphingomyelin, and bile acids) across a cellular membrane by the 67076, 67102, 44181, 67084FL, or 67084alt protein in which it is contained; and/or to regulate the activity (as defined herein) of the 67076, 67102, 44181, 67084FL, or 67084alt protein in which it is contained. Accordingly, identifying the presence of an "E1-E2 ATPases phosphorylation site" can include isolating a fragment of a 67076, 67102, 44181, 67084FL, or 67084alt molecule (*e.g.*, a 67076, 67102, 44181, 67084FL, or 67084alt polypeptide) and assaying for the ability of the fragment to exhibit one of the aforementioned phosphorylation site activities.

In another embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt protein of the present invention may also be identified based on its ability to adopt an E1 conformation or

an E2 conformation. As used herein, an "E1 conformation" of a 67076, 67102, 44181, 67084FL, or 67084alt protein includes a 3-dimensional conformation of a 67076, 67102, 44181, 67084FL, or 67084alt protein which does not exhibit 67076, 67102, 44181, 67084FL, or 67084alt activity (e.g., the ability to transport phospholipids), as defined
5 herein. An E1 conformation of a 67076, 67102, 44181, 67084FL, or 67084alt protein usually occurs when the 67076, 67102, 44181, 67084FL, or 67084alt protein is unphosphorylated. As used herein, an "E2 conformation" of a 67076, 67102, 44181, 67084FL, or 67084alt protein includes a 3-dimensional conformation of a 67076, 67102, 44181, 67084FL, or 67084alt protein which exhibits 67076, 67102, 44181, 67084FL, or
10 67084alt activity (e.g., the ability to transport phospholipids), as defined herein. An E2 conformation of a 67076, 67102, 44181, 67084FL, or 67084alt protein usually occurs when the 67076, 67102, 44181, 67084FL, or 67084alt protein is phosphorylated.

In still another embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt protein of the present invention is identified based on the presence of "phospholipid transporter
15 specific" amino acid residues. As used herein, "phospholipid transporter specific" amino acid residues are amino acid residues specific to the class of phospholipid transporting P-type ATPases (as defined in Tang, X. et al. (1996) *Science* 272:1495-1497). Phospholipid transporter specific amino acid residues are not found in those P-type ATPases which transport molecules which are not phospholipids (e.g., cations). For example, phospholipid
20 transporter specific amino acid residues are found at the first, second, and fifth positions of the P-type ATPase sequence 1 motif. In phospholipid transporting P-type ATPases, the first position of the P-type ATPase sequence 1 motif is preferably E (glutamic acid), the second position is preferably T (threonine), and the fifth position is preferably L (leucine). A phospholipid transporter specific amino acid residue is further found at the second position
25 of the P-type ATPase sequence 2 motif. In phospholipid transporting P-type ATPases, the second position of the P-type ATPase sequence 2 motif is preferably F (phenylalanine). Phospholipid transporter specific amino acid residues are still further found at the first, tenth, and eleventh positions of the P-type ATPase sequence 3 motif. In phospholipid transporting P-type ATPases, the first position of the P-type ATPase sequence 3 motif is
30 preferably I (isoleucine), the tenth position is preferably M (methionine), and the eleventh position is preferably I (isoleucine).

Phospholipid transporter specific amino acid residues were identified in the amino acid sequence of human 67076 (SEQ ID NO:14) at about residues 174 and 177 (within the P-type ATPase sequence 1 motif), at about residue 407 (within the P-type ATPase sequence
35 2 motif), and at about residues 813, 823, and 824 (within the P-type ATPase sequence 3 motif).

Phospholipid transporter specific amino acid residues were identified in the amino acid sequence of human 67102 (SEQ ID NO:17) at about residues 208, 209, and 212 (within

the P-type ATPase sequence 1 motif), at about residue 436 (within the P-type ATPase sequence 2 motif), and at about residues 1054, 1063, and 1064 (within the P-type ATPase sequence 3 motif).

5 Phospholipid transporter specific amino acid residues were identified in the amino acid sequence of human 44181 (SEQ ID NO:20) at about residues 174 and 177 (within the P-type ATPase sequence 1 motif), at about residue 405 (within the P-type ATPase sequence 2 motif), and at about residues 819, 828, and 829 (within the P-type ATPase sequence 3 motif).

10 Phospholipid transporter specific amino acid residues were identified in the amino acid sequence of human 67084FL (SEQ ID NO:23) at about residues 214 and 217 (within the P-type ATPase sequence 1 motif) and at about residues 820, 829, and 830 (within the P-type ATPase sequence 3 motif).

15 Phospholipid transporter specific amino acid residues were identified in the amino acid sequence of human 67084alt (SEQ ID NO:26) at about residues 214 and 217 (within the P-type ATPase sequence 1 motif), and at about residues 820, 829, and 830 (within the P-type ATPase sequence 3 motif).

Isolated polypeptides of the present invention, preferably 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26 or are encoded by a nucleotide sequence sufficiently identical to SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, or SEQ ID NO:27. For example, amino acid or nucleotide sequences which share common structural domains having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homology or identity across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently identical. Furthermore, amino acid or nucleotide sequences which share at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homology or identity and share a common functional activity are defined herein as sufficiently identical.

30 In a preferred embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt protein includes at least one or more of the following domains, sites, or motifs: a transmembrane domain, an N-terminal large extramembrane domain, a C-terminal large extramembrane domain, an E1-E2 ATPases phosphorylation site, a P-type ATPase sequence 1 motif, a P-type ATPase sequence 2 motif, a P-type ATPase sequence 3 motif, and/or one or more phospholipid transporter specific amino acid residues, and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous or identical to the amino acid sequence of SEQ ID NO:14, SEQ

ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____. In yet another preferred embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt protein includes at least one or more of the following domains, sites, or motifs: a transmembrane domain, an N-terminal large extramembrane domain, a C-terminal large extramembrane domain, an E1-E2 ATPases phosphorylation site, a P-type ATPase sequence 1 motif, a P-type ATPase sequence 2 motif, a P-type ATPase sequence 3 motif, and/or one or more phospholipid transporter specific amino acid residues, and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, or SEQ ID NO:27. In another preferred embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt protein includes at least one or more of the following domains, sites, or motifs: a transmembrane domain, an N-terminal large extramembrane domain, a C-terminal large extramembrane domain, an E1-E2 ATPases phosphorylation site, a P-type ATPase sequence 1 motif, a P-type ATPase sequence 2 motif, a P-type ATPase sequence 3 motif, and/or one or more phospholipid transporter specific amino acid residues, and has a 67076, 67102, 44181, 67084FL, or 67084alt activity.

As used interchangeably herein, a "phospholipid transporter activity" or a "67076, 67102, 44181, 67084FL, or 67084alt activity" includes an activity exerted or mediated by a 67076, 67102, 44181, 67084FL, or 67084alt protein, polypeptide or nucleic acid molecule on a 67076, 67102, 44181, 67084FL, or 67084alt responsive cell or on a 67076, 67102, 44181, 67084FL, or 67084alt substrate, as determined *in vivo* or *in vitro*, according to standard techniques. In one embodiment, a phospholipid transporter activity is a direct activity, such as an association with a 67076, 67102, 44181, 67084FL, or 67084alt target molecule. As used herein, a "target molecule" or "binding partner" is a molecule with which a 67076, 67102, 44181, 67084FL, or 67084alt protein binds or interacts in nature, such that 67076, 67102, 44181, 67084FL, or 67084alt-mediated function is achieved. In an exemplary embodiment, a 67076, 67102, 44181, 67084FL, or 67084alt target molecule is a 67076, 67102, 44181, 67084FL, or 67084alt substrate (e.g., a phospholipid, ATP, or a non-67076, 67102, 44181, 67084FL, or 67084alt protein). A phospholipid transporter activity can also be an indirect activity, such as a cellular signaling activity mediated by interaction of the 67076, 67102, 44181, 67084FL, or 67084alt protein with a 67076, 67102, 44181, 67084FL, or 67084alt substrate.

In a preferred embodiment, a phospholipid transporter activity is at least one of the following activities: (i) interaction with a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (e.g., a phospholipid, ATP, or a non-67076, 67102, 44181,

67084FL, or 67084alt protein); (ii) transport of a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (e.g., an aminophospholipid such as phosphatidylserine or phosphatidylethanolamine) from one side of a cellular membrane to the other; (iii) the ability to be phosphorylated or dephosphorylated; (iv) adoption of an E1 conformation or an E2 conformation; (v) conversion of a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule to a product (e.g., hydrolysis of ATP); (vi) interaction with a second non-67076, 67102, 44181, 67084FL, or 67084alt protein; (vii) modulation of substrate or target molecule location (e.g., modulation of phospholipid location within a cell and/or location with respect to a cellular membrane); (viii) maintenance of aminophospholipid gradients; (ix) modulation of intra- or intercellular signaling and/or gene transcription (e.g., either directly or indirectly); and/or (x) modulation of cellular proliferation, growth, differentiation, apoptosis, absorption, or secretion.

The nucleotide sequence of the isolated human 67076, 67102, 44181, 67084FL, or 67084alt cDNA and the predicted amino acid sequence of the human 67076, 67102, 44181, 67084FL, or 67084alt polypeptides are shown in Figures 20A-E, 24A-E, 28A-E, 32A-E, and 36A-E, and in SEQ ID NOs:13 and 14, SEQ ID NOs:16 and 17, SEQ ID NOs:19 and 20, SEQ ID NOs:22 and 23, and SEQ ID NOs:25 and 26, respectively. Plasmids containing the nucleotide sequence encoding human 67076, human 67102, human 44181, human 67084FL, and/or human 67084alt were deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, _____, _____, _____, and _____, respectively, and assigned Accession Numbers _____, _____, _____, _____, and _____, respectively. These deposits will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. These deposits were made merely as a convenience for those of skill in the art and are not admissions that a deposit is required under 35 U.S.C. §112.

The human 67076 gene, which is approximately 6582 nucleotides in length, encodes a polypeptide which is approximately 1129 amino acid residues in length. The human 67102 gene, which is approximately 6074 nucleotides in length, encodes a polypeptide which is approximately 1426 amino acid residues in length. The human 44181 gene, which is approximately 7221 nucleotides in length, encodes a polypeptide which is approximately 1177 amino acid residues in length. The human 67084FL gene, which is approximately 4198 nucleotides in length, encodes a polypeptide which is approximately 1084 amino acid residues in length. The human 67084alt gene, which is approximately 4231 nucleotides in length, encodes a polypeptide which is approximately 1095 amino acid residues in length.

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

5 One aspect of the invention pertains to isolated nucleic acid molecules that encode 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt-encoding nucleic acid molecules (*e.g.*, 8099, 46455, 54414, 53763, 10 67076, 67102, 44181, 67084FL, or 67084alt mRNA) and fragments for use as PCR primers for the amplification or mutation of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide 15 analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes 20 nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated 8099, 46455, 54414, 53763, 67076, 25 67102, 44181, 67084FL, or 67084alt nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by 30 recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ 35 ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____ or a portion thereof, can be isolated using standard

molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____, as a hybridization probe, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____ can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In one embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1. The sequence of SEQ ID NO:1 corresponds to the human 8099 cDNA. This cDNA comprises sequences encoding the human 8099 polypeptide (i.e., "the coding region", from nucleotides 180-2034) as well as 5' untranslated sequences (nucleotides 1-179) and 3' untranslated sequences (nucleotides 2035-2725). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (e.g., nucleotides 180-2034, corresponding to SEQ ID NO:3). Accordingly, in another

embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:3 and nucleotides 1-179 and 2035-2725 of SEQ ID NO:1. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:1 or SEQ ID NO:3.

5 In another embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:4. The sequence of SEQ ID NO:4 corresponds to the human 46455 cDNA. This cDNA comprises sequences encoding the human 46455 polypeptide (*i.e.*, "the coding region", from nucleotides 376-1963) as well as 5' untranslated sequences (nucleotides 1-375) and 3' untranslated sequences (nucleotides 1964-2230). Alternatively, the nucleic acid molecule can comprise only the coding region
10 of SEQ ID NO:4 (*e.g.*, nucleotides 376-1963, corresponding to SEQ ID NO:6).

Accordingly, in another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:6 and nucleotides 1-375 and 1964-2230 of SEQ ID NO:4. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:4 or SEQ ID NO:6.

15 In another embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:7. This cDNA may comprise sequences encoding the human 54414 protein (*e.g.*, the "coding region", from nucleotides 225-3578), as well as 5' untranslated sequence (nucleotides 1-224) and 3' untranslated sequences (nucleotides 3579-4632) of SEQ ID NO:7. Alternatively, the nucleic acid molecule can
20 comprise only the coding region of SEQ ID NO:7 (*e.g.*, nucleotides 225-3578, corresponding to SEQ ID NO:9). Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention comprises SEQ ID NO:9 and nucleotides 1-224 of SEQ ID NO:7. In yet another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:9 and nucleotides 3579-4632 of SEQ ID NO:7. In yet another embodiment, the nucleic
25 acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:7 or SEQ ID NO:9.

In still another embodiment, the cDNA may comprise sequences encoding the human 53763 protein (*e.g.*, the "coding region", from nucleotides 561-2474), as well as 5' untranslated sequence (nucleotides 1-560) and 3' untranslated sequences (nucleotides 2475-
30 2847) of SEQ ID NO:10. Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:10 (*e.g.*, nucleotides 561-2474, corresponding to SEQ ID NO:6). Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention comprises SEQ ID NO:12 and nucleotides 1-560 of SEQ ID NO:10. In yet another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:12 and
35 nucleotides 2475-2847 of SEQ ID NO:10. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:10 or SEQ ID NO:12.

In yet another embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:13. The sequence of SEQ ID

NO:13 corresponds to the human 67076 cDNA. This cDNA comprises sequences encoding the human 67076 polypeptide (*i.e.*, "the coding region", from nucleotides 524-3910) as well as 5' untranslated sequences (nucleotides 1-523) and 3' untranslated sequences (nucleotides 3911-6582). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:13 (*e.g.*, nucleotides 524-3910, corresponding to SEQ ID NO:15).

Accordingly, in another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:15 and nucleotides 1-523 or 3911-6582 of SEQ ID NO:13. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:13 or SEQ ID NO:15.

In another embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:16. The sequence of SEQ ID NO:16 corresponds to the human 67102 cDNA. This cDNA comprises sequences encoding the human 67102 polypeptide (*i.e.*, "the coding region", from nucleotides 274-4551) as well as 5' untranslated sequences (nucleotides 1-273) and 3' untranslated sequences (nucleotides 4552-6074). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:16 (*e.g.*, nucleotides 274-4551, corresponding to SEQ ID NO:18).

Accordingly, in another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:18 and nucleotides 1-273 or 4552-6074 of SEQ ID NO:16. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:16 or SEQ ID NO:18.

In still another embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:19. The sequence of SEQ ID NO:19 corresponds to the human 44181 cDNA. This cDNA comprises sequences encoding the human 44181 polypeptide (*i.e.*, "the coding region", from nucleotides 167-3697) as well as 5' untranslated sequences (nucleotides 1-166) and 3' untranslated sequences (nucleotides 3698-7221). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:19 (*e.g.*, nucleotides 167-3697, corresponding to SEQ ID NO:21).

Accordingly, in another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:21 and nucleotides 1-166 or 3698-7221 of SEQ ID NO:19. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:19 or SEQ ID NO:21.

In yet another embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:22. The sequence of SEQ ID NO:22 corresponds to the human 67084FL cDNA. This cDNA comprises sequences encoding the human 67084FL polypeptide (*i.e.*, "the coding region", from nucleotides 156-3407) as well as 5' untranslated sequences (nucleotides 1-155) and 3' untranslated sequences (nucleotides 3408-4198). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:22 (*e.g.*, nucleotides 156-3407, corresponding to SEQ

ID NO:24). Accordingly, in another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:24 and nucleotides 1-155 or 3408-4198 of SEQ ID NO:22. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:22 or SEQ ID NO:24.

5 In a further embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:25. The sequence of SEQ ID NO:25 corresponds to the human 67084alt cDNA. This cDNA comprises sequences encoding the human 67084alt polypeptide (*i.e.*, "the coding region", from nucleotides 156-3440) as well as 5' untranslated sequences (nucleotides 1-155) and 3' untranslated
10 sequences (nucleotides 3441-4231). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:25 (*e.g.*, nucleotides 156-3440, corresponding to SEQ ID NO:27). Accordingly, in another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:27 and nucleotides 1-155 or 3441-4231 of SEQ ID NO:25. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth
15 as SEQ ID NO:25 or SEQ ID NO:27.

In still another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16,
20 SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ
25 ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____, is one which is sufficiently
30 complementary to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number
35 _____, _____, _____, _____, or _____, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22,

SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27 (e.g., to the entire length of the nucleotide sequence), or to the nucleotide sequence (e.g., the entire length of the nucleotide sequence) of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____, or a portion of any of these nucleotide sequences.

In one embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least (or no greater than) 50-100, 100-250, 250-500, 500-750, 750-1000, 1000-1250, 1250-1500, 1500-1750, 1750-2000, 2000-2250, 2250-2500, 2500-2750, 2750-3000, 3000-3250, 3250-3500, 3500-3750, 3750-4000, 4000-4250, 4250-4500, 4500-4750, 4750-5000, 5000-5250, 5250-5500, 5500-5750, 5750-6000, 6000-6250, 6250-6500, 6500-6750, 6750-7000, 7000-7250, 7250-7500 or more nucleotides in length and hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____, for example, a fragment which can be used as a probe or primer or a fragment encoding a portion of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, e.g., a biologically active portion of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. The nucleotide sequence determined from the cloning of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 8099, 46455, 54414, 53763, 67076,

67102, 44181, 67084FL, or 67084alt family members, as well as 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The probe/primer (e.g., oligonucleotide) typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, 75, 80, 85, 90, 95, or 100 or more consecutive nucleotides of a sense sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____, of an anti-sense sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27 or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____.

Exemplary probes or primers are at least 12, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75 or more nucleotides in length and/or comprise consecutive nucleotides of an isolated nucleic acid molecule described herein. Probes based on the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences can be used to detect (e.g., specifically detect) transcripts or genomic sequences encoding the same or homologous polypeptides. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. In another embodiment a set of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence, e.g., a domain, region, site or other sequence described herein. The primers should be at least 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more nucleotides in length. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, such as by measuring a level of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or

67084alt -encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA levels or determining whether a genomic 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene has been mutated or deleted.

5 A nucleic acid fragment encoding a "biologically active portion of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID
10 NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____, which encodes a polypeptide having a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt biological activity (the biological activities of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides are described herein), expressing the encoded portion of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide (*e.g.*, by
15 recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. In an exemplary embodiment, the nucleic acid molecule is at least 50-100, 100-250, 250-500, 500-750, 750-1000, 1000-1250, 1250-1500, 1500-1750, 1750-2000, 2000-2250, 2250-2500, 2500-2750, 2750-3000, 3000-3250, 3250-3500, 3500-3750, 3750-4000, 4000-4250, 4250-4500, 4500-4750, 4750-5000, 5000-5250, 5250-5500, 5500-5750, 5750-6000, 6000-6250, 6250-6500, 6500-6750, 6750-7000, 7000-7250, 7250-7500 or more nucleotides in length and encodes a polypeptide having a 8099, 46455, 54414, 53763, 67076, 67102, 44181,
20 67084FL, or 67084alt activity (as described herein).

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA
30 insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____. Such differences can be due to degeneracy of the genetic code, thus resulting in a nucleic acid which encodes the same 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides as those encoded by the
35 nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA

insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a polypeptide having an amino acid sequence which differs by at least 1, but no greater than 5, 10, 20, 50 or 100 amino acid residues from the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____, _____, _____, or _____. In yet another embodiment, the nucleic acid molecule encodes the amino acid sequence of human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. If an alignment is needed for this comparison, the sequences should be aligned for maximum homology.

Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologues (different locus), and orthologues (different organism) or can be non naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

Allelic variants result, for example, from DNA sequence polymorphisms within a population (e.g., the human population) that lead to changes in the amino acid sequences of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides. Such genetic polymorphism in the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, preferably a mammalian 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, and can further include non-coding regulatory sequences, and introns.

Accordingly, in one embodiment, the invention features isolated nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, or _____, wherein the nucleic acid molecule hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18,

SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, for example, under stringent hybridization conditions.

Allelic variants of human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt include both functional and non-functional 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides.

Functional allelic variants are naturally occurring amino acid sequence variants of the human 8099 or 46455 polypeptides that have an 8099 or 46455 activity, *e.g.*, maintain the ability to bind an 8099 or 46455 ligand or substrate and/or modulate sugar transport, or sugar homeostasis.

Functional allelic variants are naturally occurring amino acid sequence variants of the human 54414 or 53763 polypeptides that maintain the ability to, *e.g.*, bind or interact with a 54414 or 53763 target molecule and/or modulate membrane excitability.

Functional allelic variants are naturally occurring amino acid sequence variants of the human 67076, 67102, 44181, 67084FL, or 67084alt polypeptides that have a 67076, 67102, 44181, 67084FL, or 67084alt activity, *e.g.*, bind or interact with a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule, transport a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule across a cellular membrane, hydrolyze ATP, be phosphorylated or dephosphorylated, adopt an E1 conformation or an E2 conformation, and/or modulate cellular signaling, growth, proliferation, differentiation, absorption, or secretion.

Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, or substitution, deletion or insertion of non-critical residues in non-critical regions of the polypeptide.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the human 8099 or 46455 polypeptides that do not have a 8099 or 46455 activity, *e.g.*, maintain the ability to bind an 8099 or 46455 ligand or substrate and/or modulate sugar transport, or sugar homeostasis.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the human 54414 or 53763 polypeptides that do not maintain the ability to, *e.g.*, bind or interact with a 54414 or 53763 target molecule and/or modulate membrane excitability.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the human 67076, 67102, 44181, 67084FL, or 67084alt polypeptides that do not have a 67076, 67102, 44181, 67084FL, or 67084alt activity, *e.g.*, that do not have the ability to, *e.g.*, bind or interact with a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule, transport a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule across a cellular membrane, hydrolyze ATP, be phosphorylated or

dephosphorylated, adopt an E1 conformation or an E2 conformation, and/or modulate cellular signaling, growth, proliferation, differentiation, absorption, or secretion.

Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion or premature truncation of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, or a substitution, insertion or deletion in critical residues or critical regions.

The present invention further provides non-human orthologues of the human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides.

10. Orthologues of human 8099 or 46455 polypeptides are polypeptides that are isolated from non-human organisms and possess the same 8099 and/or 46455 activity, *e.g.*, ligand binding and/or modulation of sugar transport mechanisms, as the human 8099 and/or 46455 polypeptide. Orthologues of the human 54414 or 53763 polypeptides are polypeptides that are isolated from non-human organisms and possess the same 54414 or 53763 target molecule binding mechanisms and/or ability to modulate membrane excitability of the human 54414 or 53763 polypeptides. Orthologues of human 67076, 67102, 44181, 67084FL, or 67084alt polypeptides are polypeptides that are isolated from non-human organisms and possess the same 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule binding mechanisms, phospholipid transporting activity, ATPase activity, and/or modulation of cellular signaling mechanisms of the human 67076, 67102, 44181, 67084FL, or 67084alt proteins as the human 67076, 67102, 44181, 67084FL, or 67084alt polypeptides.

Orthologues of the human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides can readily be identified as comprising an amino acid sequence that is substantially identical to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26.

Moreover, nucleic acid molecules encoding other 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt family members and, thus, which have a nucleotide sequence which differs from the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____ are intended to be within the scope of the invention. For example, another 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt cDNA can be identified based on the nucleotide sequence of human

8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. Moreover, nucleic acid molecules encoding 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides from different species, and which, thus, have a nucleotide sequence which differs from the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____ are intended to be within the scope of the invention. For example, a mouse 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt cDNA can be identified based on the nucleotide sequence of a human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt cDNAs of the invention can be isolated based on their homology to the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Nucleic acid molecules corresponding to natural allelic variants and homologues of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene.

Orthologues, homologues and allelic variants can be identified using methods known in the art (e.g., by hybridization to an isolated nucleic acid molecule of the present invention, for example, under stringent hybridization conditions). In one embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____. In other embodiment, the nucleic acid is at least 50-100, 100-250, 250-500, 500-750, 750-1000, 1000-1250, 1250-1500, 1500-1750, 1750-2000, 2000-2250, 2250-2500, 2500-2750, 2750-3000, 3000-3250, 3250-3500, 3500-3750, 3750-4000, 4000-4250, 4250-4500, 4500-4750, 4750-5000, 5000-5250, 5250-5500, 5500-5750,

5750-6000, 6000-6250, 6250-6500, 6500-6750, 6750-7000, 7000-7250, 7250-7500 or more nucleotides in length.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences that are significantly identical or homologous to each other remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, Inc. (1995), sections 2, 4 and 6. Additional stringent conditions can be found in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), chapters 7, 9 and 11. A preferred, non-limiting example of stringent hybridization conditions includes hybridization in 4X sodium chloride/sodium citrate (SSC), at about 65-70°C (or hybridization in 4X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 1X SSC, at about 65-70°C. A preferred, non-limiting example of highly stringent hybridization conditions includes hybridization in 1X SSC, at about 65-70°C (or hybridization in 1X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 0.3X SSC, at about 65-70°C. A preferred, non-limiting example of reduced stringency hybridization conditions includes hybridization in 4X SSC, at about 50-60°C (or alternatively hybridization in 6X SSC plus 50% formamide at about 40-45°C) followed by one or more washes in 2X SSC, at about 50-60°C. Ranges intermediate to the above-recited values, *e.g.*, at 65-70°C or at 42-50°C are also intended to be encompassed by the present invention. SSPE (1xSSPE is 0.15M NaCl, 10mM NaH₂PO₄, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes each after hybridization is complete. The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, $T_m(^{\circ}\text{C}) = 2(\# \text{ of A + T bases}) + 4(\# \text{ of G + C bases})$. For hybrids between 18 and 49 base pairs in length, $T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\% \text{G+C}) - (600/\text{N})$, where N is the number of bases in the hybrid, and $[\text{Na}^+]$ is the concentration of sodium ions in the hybridization buffer ($[\text{Na}^+]$ for 1xSSC = 0.165 M). It will also be recognized by the skilled practitioner that additional reagents may be added to hybridization and/or wash buffers to decrease non-specific hybridization of nucleic acid molecules to membranes, for example, nitrocellulose or nylon membranes, including but not limited to blocking agents (*e.g.*, BSA or salmon or herring sperm carrier DNA), detergents (*e.g.*, SDS), chelating agents (*e.g.*, EDTA), Ficoll, PVP and the like. When using nylon membranes, in particular, an additional

preferred, non-limiting example of stringent hybridization conditions is hybridization in 0.25-0.5M NaH_2PO_4 , 7% SDS at about 65°C, followed by one or more washes at 0.02M NaH_2PO_4 , 1% SDS at 65°C, see e.g., Church and Gilbert (1984) *Proc. Natl. Acad. Sci. USA* 81:1991-1995, (or alternatively 0.2X SSC, 1% SDS).

5 Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, and corresponds to a naturally-
10 occurring nucleic acid

molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural polypeptide).

In addition to naturally-occurring allelic variants of the 8099, 46455, 54414, 53763, 15 67076, 67102, 44181, 67084FL, or 67084alt sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID
20 NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____, thereby leading to changes in the amino acid sequence of the encoded 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, without altering the functional ability of the 8099, 46455, 54414, 53763, 67076, 67102, 44181,
25 67084FL, or 67084alt polypeptides. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25,
30 SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt (e.g., the sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID
35 NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity.

For example, amino acid residues that are conserved among the 8099 or 46455 polypeptides of the present invention, e.g., those present in a transmembrane domain and/or a

sugar transporter family domain, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the 8099 or 46455 polypeptides of the present invention, and other members of the 8099 or 46455 family are not likely to be amenable to alteration.

5 Amino acid residues that are conserved among the 54414 or 53763 polypeptides of the present invention, *e.g.*, those present in a P-loop or a pore domain, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the 54414 or 53763 polypeptides of the present invention and other members of the potassium channel family are not likely to be amenable to alteration.

10 Amino acid residues that are conserved among the 67076, 67102, 44181, 67084FL, or 67084alt polypeptides of the present invention, *e.g.*, those present in a E1-E2 ATPases phosphorylation site, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the 67076, 67102, 44181, 67084FL, or 67084alt polypeptides of the present invention and other members of the
15 phospholipid transporter family are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides that contain changes in amino acid residues that are not essential for activity. Such 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides
20 differ in amino acid sequence from SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%,
25 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26 (*e.g.*, to the entire length of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26).

30 An isolated nucleic acid molecule encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide identical to the polypeptide of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1,
35 SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession

Number _____, _____, _____, _____, or _____, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded polypeptide. Mutations can be introduced into SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____ by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine, tryptophan), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, _____, _____, _____, _____, or _____, the encoded polypeptide can be expressed recombinantly and the activity of the polypeptide can be determined.

In a preferred embodiment, a mutant 8099 and/or 46455 polypeptide can be assayed for the ability to (1) bind a monosaccharide, *e.g.*, D-glucose, D-fructose, D-galactose, and/or mannose, (2) transport monosaccharides across a cell membrane, (3) influence insulin and/or glucagon secretion, (4) maintain sugar homeostasis in a cell, (5) function as a neuronal transporter, and (6) mediate trans-epithelial movement in a cell.

In another preferred embodiment, a mutant 54414 and/or 53763 protein can be assayed for the ability to (i) interact with a 54414 and/or 53763 substrate (e.g., a potassium ion or a cyclic nucleotide); (ii) conduct or transport a 54414 and/or 53763 substrate across a cellular membrane; (iii) interact with a second non-54414 and/or 53763 protein (e.g., a 54414 and/or 53763 polypeptide or a 54414 and/or 53763 -potassium channel subunit); (iv) modulate (e.g., maintain and/or rectify) membrane potentials; (v) regulate target molecule availability or activity; (vi) modulate intra- or intercellular signaling and/or gene transcription (e.g., either directly or indirectly); (viii) generate outwardly rectifying currents; (viii) modulate membrane excitability; (ix) modulate the release of neurotransmitters; (x) regulate contractility (e.g., of smooth muscle cells), secretion, and/or synaptic transmission; and/or (xi) modulate processes which underlie learning and memory.

In a further preferred embodiment, a mutant 54414 protein can be assayed for the ability to (i) interact with maxi-K potassium channels (i.e., large conductance channels, in particular *Slo*); (ii) modulate maxi-K potassium channel activity (e.g., *Slo*-mediated activities); (iii) generate intermediate conductance channels; and/or (iv) regulate contractility (e.g., of smooth muscle cells), secretion, and/or synaptic transmission, in particular, via modulation of *Slo*.

In still a further preferred embodiment, a mutant 53763 protein can be assayed for the ability to (i) interact with Shaker (Sh) potassium channels and/or channel subunits; (ii) modulate Shaker (Sh) potassium channel activity (e.g., termination of prolonged membrane depolarization); and/or (iii) modulation of high voltage activating channel activity and/or inactivating channel activity, and the like.

In yet another preferred embodiment, a mutant 67076, 67102, 44181, 67084FL, and/or 67084alt polypeptide can be assayed for the ability to (i) interact with a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (e.g., a phospholipid, ATP, or a non-67076, 67102, 44181, 67084FL, or 67084alt protein); (ii) transport a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (e.g., an aminophospholipid such as phosphatidylserine or phosphatidylethanolamine) from one side of a cellular membrane to the other; (iii) be phosphorylated or dephosphorylated; (iv) adopt an E1 conformation or an E2 conformation; (v) convert a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule to a product (e.g., hydrolysis of ATP); (vi) interact with a second non-67076, 67102, 44181, 67084FL, or 67084alt protein; (vii) modulate substrate or target molecule location (e.g., modulation of phospholipid location within a cell and/or location with respect to a cellular membrane); (viii) maintain aminophospholipid gradients; (ix) modulate intra- or intercellular signaling and/or gene transcription (e.g., either directly or indirectly); and/or (x) modulate cellular proliferation, growth, differentiation, apoptosis, absorption, or secretion.

In addition to the nucleic acid molecules encoding 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. In an exemplary embodiment, the invention provides an isolated nucleic acid molecule which is antisense to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecule (e.g., is antisense to the coding strand of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecule). An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a polypeptide, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the coding region of human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt corresponds to SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:9, SEQ ID NO:12, and SEQ ID NO:15, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:24, and SEQ ID NO:27, respectively). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt disclosed herein (e.g., SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:9, SEQ ID NO:12, and SEQ ID NO:15, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:24, and SEQ ID NO:27), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA (e.g., between the -10 and +10 regions of the start site of a gene nucleotide sequence). An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50

nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide to thereby inhibit expression of the polypeptide, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve

sufficient intra-cellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-*o*-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA transcripts to thereby inhibit translation of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA. A ribozyme having specificity for a 8099-, 46455-, 54414-, 53763-, 67076-, 67102-, 44181-, 67084FL-, or 67084alt -encoding nucleic acid can be designed based upon the nucleotide sequence of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt cDNA disclosed herein (*i.e.*, SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 8099-, 46455-, 54414-, 53763-, 67076-, 67102-, 44181-, 67084FL-, or 67084alt -encoding mRNA. See, e.g., Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt (e.g., the 8099, 46455, 54414, 53763, 67076, 67102, 44181,

67084FL, or 67084alt promoter and/or enhancers) to form triple helical structures that prevent transcription of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al. Proc. Natl. Acad. Sci.* 93: 14670-675.

PNAs of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can be modified, (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (*e.g.*, RNase H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of

bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl) amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

Alternatively, the expression characteristics of an endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene within a cell line or microorganism may be modified by inserting a heterologous DNA regulatory element into the genome of a stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene. For example, an endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene which is normally "transcriptionally silent", *i.e.*, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene which is normally not expressed, or is expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, a transcriptionally silent, endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene may be activated by insertion of a promiscuous regulatory element that works across cell types.

A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with an endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, using techniques, such as

targeted homologous recombination, which are well known to those of skill in the art, and described, e.g., in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

5 II. Isolated 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt Polypeptides and Anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt Antibodies

One aspect of the invention pertains to isolated 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt or recombinant polypeptides and polypeptides, and
10 biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibodies. In one embodiment, native 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another
15 embodiment, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides are produced by recombinant DNA techniques. Alternative to recombinant expression, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

20 An "isolated" or "purified" polypeptide or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular
25 material" includes preparations of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide in which the polypeptide is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide
30 having less than about 30% (by dry weight) of non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, still more preferably less than about 10% of non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or
35 67084alt polypeptide, and most preferably less than about 5% non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. When the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of

culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes
5 preparations of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide in which the polypeptide is separated from chemical precursors or other chemicals which are involved in the synthesis of the polypeptide. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes
10 preparations of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide having less than about 30% (by dry weight) of chemical precursors or non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt chemicals, more preferably less than about 20% chemical precursors or non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt chemicals, still more preferably less than about
15 10% chemical precursors or non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt chemicals, and most preferably less than about 5% chemical precursors or non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt chemicals.

As used herein, a "biologically active portion" of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide includes a fragment of a 8099,
20 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide which participates in an interaction between a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecule and a non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecule (*e.g.*, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate). Biologically active portions of a 8099, 46455,
25 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, *e.g.*, the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or
30 SEQ ID NO:26, which include less amino acids than the full length 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, and exhibit at least one activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide.

Typically, biologically active portions of a 8099 or 46455 polypeptide comprise a
35 domain or motif with at least one activity of the 8099 or 46455 polypeptide, *e.g.*, modulating sugar transport mechanisms. A biologically active portion of an 8099 polypeptide can be a polypeptide which is, for example, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 600 or more amino acids in

length. A biologically active portion of an 46455 polypeptide can be a polypeptide which is, for example, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525 or more amino acids in length. Biologically active portions of an 8099 and/or an 46455 polypeptide can be used as targets for developing agents which modulate an 8099 or 46455 mediated activity, e.g., a sugar transport mechanism.

In one embodiment, a biologically active portion of an 8099 or an 46455 polypeptide comprises at least one transmembrane domain. It is to be understood that a preferred biologically active portion of an 8099 or an 46455 polypeptide of the present invention comprises at least one or more of the following domains: a transmembrane domain and/or a sugar transporter family domain. Moreover, other biologically active portions, in which other regions of the polypeptide are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 8099 or 46455 polypeptide.

Moreover, biologically active portions of a 54414 or 53763 polypeptide comprise a domain or motif with at least one activity of the 54414 or 53763 polypeptide, e.g., modulation of intra- or inter-cellular signaling and/or gene expression, and/or modulate membrane excitability. A biologically active portion of a 54414 or 53763 polypeptide can be a polypeptide which is, for example, 10, 25, 50, 75, 100, 125, 150 or more amino acids in length. Biologically active portions of a 54414 or 53763 polypeptide can be used as targets for developing agents which modulate a 54414 or 53763 mediated activity, e.g., modulation of intra- or inter-cellular signaling and/or gene expression, and/or modulate membrane excitability.

In one embodiment, a biologically active portion of a 54414 or 53763 polypeptide comprises at least one transmembrane domain and/or a pore domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 54414 or 53763 polypeptide.

Biologically active portions of a 67076, 67102, 44181, 67084FL, or 67084alt polypeptide comprise a domain or motif with at least one activity of the 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, e.g., the ability to interact with a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (e.g., a phospholipid; ATP; a non-67076, 67102, 44181, 67084FL, or 67084alt protein; or another 67076, 67102, 44181, 67084FL, or 67084alt protein or subunit); the ability to transport a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (e.g., a phospholipid) from one side of a cellular membrane to the other; the ability to be phosphorylated or dephosphorylated; the ability to adopt an E1 conformation or an E2 conformation; the ability to convert a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule to a product (e.g., the ability to hydrolyze ATP); the ability to interact with a second non-67076, 67102, 44181,

67084FL, or 67084alt protein; the ability to modulate intra- or inter-cellular signaling and/or gene transcription (e.g., either directly or indirectly); the ability to modulate cellular growth, proliferation, differentiation, absorption, and/or secretion. A biologically active portion of a 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can be a polypeptide which is, for example, 10, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550 or more amino acids in length. Biologically active portions of a 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can be used as targets for developing agents which modulate a 67076, 67102, 44181, 67084FL, or 67084alt mediated activity, e.g., modulating transport of biological molecules across membranes.

In one embodiment, a biologically active portion of a 67076, 67102, 44181, 67084FL, or 67084alt polypeptide comprises at least one at least one or more of the following domains, sites, or motifs: a transmembrane domain, an N-terminal large extramembrane domain, a C-terminal large extramembrane domain, an E1-E2 ATPases phosphorylation site, a P-type ATPase sequence 1 motif, a P-type ATPase sequence 2 motif, a P-type ATPase sequence 3 motif, and/or one or more phospholipid transporter specific amino acid residues. Moreover, other biologically active portions, in which other regions of the polypeptide are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 67076, 67102, 44181, 67084FL, or 67084alt polypeptide.

Another aspect of the invention features fragments of the polypeptide having the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, for example, for use as immunogens. In one embodiment, a fragment comprises at least 5 amino acids (e.g., contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____, _____, _____, _____, or _____. In another embodiment, a fragment comprises at least 10, 15, 20, 25, 30, 35, 40, 45, 50 or more amino acids (e.g., contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____, _____, _____, _____, or _____.

In a preferred embodiment, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide has an amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID

NO:20, SEQ ID NO:23, or SEQ ID NO:26. In other embodiments, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide is substantially identical to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26; and retains the functional activity of the polypeptide of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. In another embodiment, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide is a polypeptide which comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26.

In another embodiment, the invention features a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to a nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or a complement thereof. This invention further features a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or a complement thereof.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the 8099 amino acid sequence of SEQ ID NO:2 having 617 amino acid residues, at least 185,

preferably at least 246, more preferably at least 308, more preferably at least 370, even more preferably at least 431, and even more preferably at least 493 or 555 or more amino acid residues are aligned. In another preferred embodiment, the sequences being aligned for comparison purposes are globally aligned and percent identity is determined over the entire length of the sequences aligned. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at the Accelrys website), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package, using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A preferred, non-limiting example of parameters to be used in conjunction with the GAP program include a Blosum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*Comput. Appl. Biosci.*, 4:11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0 or version 2.0U), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and polypeptide sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules of the

invention. BLAST protein searches can be performed with the XBLAST program, score = 100, wordlength = 3, and a Blosum62 matrix to obtain amino acid sequences homologous to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt, polypeptide molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped
5 BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See the National Center for Biotechnology website.

The invention also provides 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt, chimeric or fusion proteins. As used herein, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt "chimeric protein" or "fusion protein" comprises a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide operatively linked to a non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. A "8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide whereas a "non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a polypeptide which is not substantially homologous to the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, respectively, *e.g.*, a polypeptide which is different from the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide and which is derived from the same or a different organism. Within a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt fusion protein the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can
25 correspond to all or a portion of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. In a preferred embodiment, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt fusion protein comprises at least one biologically active portion of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. In another preferred embodiment, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt fusion protein comprises at least two biologically active portions of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. Within the fusion protein, the term "operatively linked" is intended to indicate that the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide and the non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide are fused in-frame to each other. The non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can be fused to the
35 N-terminus or C-terminus of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide.

For example, in one embodiment, the fusion protein is a GST-8099, -46455, -54414, -53763, -67076, -67102, -44181, -67084FL, or -67084alt fusion protein in which the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt.

In another embodiment, the fusion protein is a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can be increased through the use of a heterologous signal sequence.

The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt fusion proteins can be used to affect the bioavailability of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate. Use of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide; (ii) mis-regulation of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene; and (iii) aberrant post-translational modification of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide.

Moreover, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -fusion proteins of the invention can be used as immunogens to produce anti-8099, anti-46455, anti-54414, anti-53763, anti-67076, anti-67102, anti-44181, anti-67084FL, and/or anti-67084alt antibodies in a subject, to purify 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt ligands and in screening assays to identify molecules which inhibit the interaction with or transport of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate.

Preferably, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques

including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide.

The present invention also pertains to variants of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides which function as either 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt agonists (mimetics) or as 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antagonists. Variants of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides can be generated by mutagenesis, e.g., discrete point mutation or truncation of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. An agonist of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. An antagonist of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can inhibit one or more of the activities of the naturally occurring form of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide by, for example, competitively modulating a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -mediated activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the polypeptide has fewer side effects in a subject relative to treatment with the naturally occurring form of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide.

In one embodiment, variants of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide which function as either 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt agonists (mimetics) or as 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide for 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide agonist or

antagonist activity. In one embodiment, a variegated library of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences therein. There are a variety of methods which can be used to produce libraries of potential 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide coding sequence can be used to generate a variegated population of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt fragments for screening and subsequent selection of variants of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of 8099, 46455, 54414,

53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt library. For example, a library of expression vectors can be transfected into a cell line, which ordinarily responds to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt in a particular substrate-dependent manner. The transfected cells are then contacted with 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt and the effect of the expression of the mutant on signaling by the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate can be detected, e.g., phospholipid transport (e.g., by measuring phospholipid levels inside the cell or its various cellular compartments, within various cellular membranes, or in the extra-cellular medium), hydrolysis of ATP, phosphorylation or dephosphorylation of the HEAT protein, and/or gene transcription. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the HEAT substrate, or which score for increased or decreased levels of phospholipid transport or ATP hydrolysis, and the individual clones further characterized.

An isolated 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt using standard techniques for polyclonal and monoclonal antibody preparation. A full-length 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can be used or, alternatively, the invention provides antigenic peptide fragments of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt for use as immunogens. The antigenic peptide of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26 and encompasses an epitope of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt such

that an antibody raised against the peptide forms a specific immune complex with 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt that are located on the surface of the polypeptide, *e.g.*, hydrophilic regions, as well as regions with high antigenicity (see, for example, Figures 2, 9, 13, 17, 21, 25, 29, 33, and 37).

10 A 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or a chemically synthesized 8099, 15 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt preparation induces a polyclonal anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 20 67084FL, or 67084alt antibody response.

Accordingly, another aspect of the invention pertains to polyclonal anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site 25 which specifically binds (immunoreacts with) an antigen, such as 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind 8099, 46455, 54414, 53763, 67076, 67102, 30 44181, 67084FL, or 67084alt. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. A monoclonal antibody composition thus typically displays a single binding affinity for a particular 8099, 35 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide with which it immunoreacts.

Polyclonal anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibodies can be prepared as described above by immunizing a suitable subject

with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt immunogen. The anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. If desired, the antibody molecules directed against 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem.* 255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt monoclonal antibody (see, e.g., G. Galfre *et al.* (1977) *Nature* 266:55052; Gefter *et al.* *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture

medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines.

These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt to thereby isolate immunoglobulin library members that bind 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrard *et al.* (1991) *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al.* *Nature* (1990) 348:552-554.

Additionally, recombinant anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and

humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; 5 Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt 15 antibody (e.g., monoclonal antibody) can be used to isolate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibody can facilitate the purification of natural 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt from cells and of 20 recombinantly produced 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expressed in host cells. Moreover, an anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibody can be used to detect 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the 8099, 25 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. Anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. 30 Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials 35 include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include

luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

III. Recombinant Expression Vectors and Host Cells

5 Another aspect of the invention pertains to vectors, for example recombinant expression vectors, containing a nucleic acid containing a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecule or vectors containing a nucleic acid molecule which encodes a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide

sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, mutant forms of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, fusion proteins, and the like).

Accordingly, an exemplary embodiment provides a method for producing a polypeptide, preferably a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, by culturing in a suitable medium a host cell of the invention (e.g., a mammalian host cell such as a non-human mammalian cell) containing a recombinant expression vector, such that the polypeptide is produced.

The recombinant expression vectors of the invention can be designed for expression of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides in prokaryotic or eukaryotic cells. For example, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5

(Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity assays, (e.g., direct assays or competitive
5 assays described in detail below), or to generate antibodies specific for 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, for example. In a preferred embodiment, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into
10 irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990)
15 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene
20 under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the
25 nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (Invitrogen Corp, San Diego, CA).
30

Alternatively, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9
35

cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine *hox* promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are

produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1)-1986.

Another aspect of the invention pertains to host cells into which a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecule of the invention is introduced, *e.g.*, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecule within a vector (*e.g.*, a recombinant expression vector) or a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or

67084alt polypeptide or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

5 A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide. Accordingly, the invention further provides methods for producing a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector
10 encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide has been introduced) in a suitable medium such that a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide is produced. In another embodiment, the method further comprises isolating a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide from the medium or the host cell.

15 The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous 8099, 46455, 54414,
20 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences have been introduced into their genome or homologous recombinant animals in which endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences have been altered. Such animals are useful for studying the function and/or activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt and for identifying and/or evaluating
25 modulators of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene
30 is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous 8099,
35 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt cDNA sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, or SEQ ID NO:13 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, such as a mouse or rat 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, can be used as a transgene. Alternatively, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene homologue, such as another 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt family member, can be isolated based on hybridization to the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt cDNA sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, or _____ (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt transgene to direct expression of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt transgene in its genome and/or expression of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene. The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:9, SEQ ID NO:12, or SEQ ID NO:15), but more preferably, is a non-human homologue of a human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene (*e.g.*, a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, or SEQ ID NO:13). For example, a mouse 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene can be used to construct a homologous recombination nucleic acid molecule, *e.g.*, a vector, suitable for altering an endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene in the mouse genome. In a preferred embodiment, the homologous recombination nucleic acid molecule is designed such that, upon homologous recombination, the endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the homologous recombination nucleic acid molecule can be designed such that, upon homologous recombination, the endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene is mutated or otherwise altered but still encodes functional polypeptide (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide). In the homologous recombination nucleic acid molecule, the altered portion of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene to allow for homologous recombination to occur between the exogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene carried by the homologous recombination nucleic acid molecule and an endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene in a cell, *e.g.*, an embryonic stem cell. The additional flanking 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the homologous recombination nucleic acid molecule (*see, e.g.*, Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The homologous recombination nucleic acid molecule is introduced into a cell, *e.g.*, an embryonic stem cell line (*e.g.*, by electroporation) and cells in

which the introduced 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene has homologously recombined with the endogenous 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene are selected (see e.g., Li, E. *et al.* (1992) *Cell* 69:915). The selected cells can then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination nucleic acid molecules, e.g., vectors, or homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules, fragments of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibodies, and or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt modulators, (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, polypeptide, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol,

propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various
5 antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum
10 monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a fragment of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or an anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibody) in the required amount in an appropriate solvent with one
15 or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-
20 drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral
therapeutic administration, the active compound can be incorporated with excipients and
25 used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or
30 compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

35 For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such

compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

As defined herein, a therapeutically effective amount of polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a polypeptide or antibody can include a single treatment or, preferably, can include a series of treatments.

In a preferred example, a subject is treated with antibody or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e.*, including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or

inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologues thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly

actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, e.g., Arnon *et al.*, "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in *Monoclonal Antibodies And Cancer Therapy*, Reisfeld *et al.* (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom *et al.*, "Antibodies For Drug Delivery", in *Controlled Drug Delivery* (2nd Ed.), Robinson *et al.* (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in *Monoclonal Antibodies '84: Biological And Clinical Applications*, Pinchera *et al.* (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in *Monoclonal Antibodies For Cancer Detection And Therapy*, Baldwin *et al.* (eds.), pp. 303-16 (Academic Press 1985), and Thorpe *et al.*, "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", *Immunol. Rev.*, 62:119-58 (1982). Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, antibodies, and modulators described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic).

As described herein, an 8099 and/or 46455 polypeptide of the invention has one or more of the following activities: (1) bind a monosaccharide, e.g., D-glucose, D-fructose, D-galactose, and/or mannose, (2) transport monosaccharides across a cell membrane, (3) influence insulin and/or glucagon secretion, (4) maintain sugar homeostasis in a cell, (5) function as a neuronal transporter, and (6) mediate trans-epithelial movement in a cell.

As described herein, a 54414 and/or 53763 protein of the invention has one or more of the following activities: (i) interaction with a 54414 or 53763 substrate (e.g., a potassium ion or a cyclic nucleotide); (ii) conductance or transport of a 54414 or 53763 substrate across a cellular membrane; (iii) interaction with a second non-54414 or 53763 protein (e.g., a 54414 or 53763 polypeptide or a non-54414 or 53763 potassium channel subunit); (iv) modulation (e.g., maintenance and/or rectification) of membrane potentials; (v) regulation of target molecule availability or activity; (vi) modulation of intra- or intercellular signaling and/or gene transcription (e.g., either directly or indirectly); (viii) generation of outwardly rectifying currents; (viii) modulation of membrane excitability; (ix) modulation of the release of neurotransmitters; (x) regulation of contractility (e.g., of smooth muscle cells), secretion, and/or synaptic transmission; and/or (xi) modulation of processes which underlie learning and memory.

Preferred activities of 54414 further include at least one of the following activities: (i) interaction with maxi-K potassium channels (i.e., large conductance channels, in particular *Slo*); (ii) modulation of maxi-K potassium channel activity (e.g., *Slo*-mediated activities); (iii) generation of intermediate conductance channels; and/or (iv) regulation of contractility (e.g., of smooth muscle cells), secretion, and/or synaptic transmission, in particular, via modulation of *Slo*.

Preferred activities of 53763 further include at least one of the following activities: (i) interaction with Shaker (Sh) potassium channels and/or channel subunits; (ii) modulation of Shaker (Sh) potassium channel activity (e.g., termination of prolonged membrane depolarization); (iii) modulation of high voltage activating channel activity and/or inactivating channel activity, and the like.

As described herein, a 67076, 67102, 44181, 67084FL, or 67084alt polypeptide of the invention has one or more of the following activities: (i) interaction with a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (e.g., a phospholipid, ATP, or a non-67076, 67102, 44181, 67084FL, or 67084alt protein); (ii) transport of a

67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (e.g., an aminophospholipid such as phosphatidylserine or phosphatidylethanolamine) from one side of a cellular membrane to the other; (iii) the ability to be phosphorylated or dephosphorylated; (iv) adoption of an E1 conformation or an E2 conformation; (v) conversion of a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule to a product (e.g., hydrolysis of ATP); (vi) interaction with a second non- 67076, 67102, 44181, 67084FL, or 67084alt protein; (vii) modulation of substrate or target molecule location (e.g., modulation of phospholipid location within a cell and/or location with respect to a cellular membrane); (viii) maintenance of aminophospholipid gradients; (ix) modulation of intra- or intercellular signaling and/or gene transcription (e.g., either directly or indirectly); and/or (x) modulation of cellular proliferation, growth, differentiation, apoptosis, absorption, or secretion.

The isolated nucleic acid molecules of the invention can be used, for example, to express 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA (e.g., in a biological sample) or a genetic alteration in a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, and to modulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity, as described further below. The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides can be used to treat disorders characterized by insufficient or excessive production of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate or production or transport of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt inhibitors, for example, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt associated disorders.

As used herein, a "sugar transporter" includes a protein or polypeptide which is involved in transporting a molecule, e.g., a monosaccharide such as D-glucose, D-fructose, D-galactose or mannose, across the plasma membrane of a cell, e.g., a liver cell, fat cell, muscle cell, or blood cell, such as an erythrocyte. Sugar transporters regulate sugar homeostasis in a cell and, typically, have sugar substrate specificity. Examples of sugar transporters include glucose transporters, fructose transporters, and galactose transporters.

As used herein, a "sugar transporter mediated activity" includes an activity which involves a sugar transporter, e.g., a sugar transporter in a liver cell, fat cell, muscle cell, or blood cell, such as an erythrocyte. Sugar transporter mediated activities include the transport of sugars, e.g., D-glucose, D-fructose, D-galactose or mannose, into and out of cells; the stimulation of molecules that regulate glucose homeostasis (e.g., insulin and glucagon), from cells, e.g., pancreatic cells; and the participation in signal transduction pathways associated with sugar metabolism.

As the 8099 and 46455 molecules of the present invention are sugar transporters, they may be useful for developing novel diagnostic and therapeutic agents for sugar transporter associated disorders. As used herein, the terms "sugar transporter associated disorder" and "8099 and 46455 disorder," used interchangeably herein, includes a disorder, disease, or condition which is characterized by an aberrant, *e.g.*, upregulated or downregulated, sugar transporter mediated activity. Sugar transporter associated disorders typically result in, *e.g.*, upregulated or downregulated, sugar levels in a cell. Examples of sugar transporter associated disorders include disorders associated with sugar homeostasis, such as obesity, anorexia, type-1 diabetes, type-2 diabetes, hypoglycemia, glycogen storage disease (Von Gierke disease), type I glycogenosis, bipolar disorder, seasonal affective disorder, and cluster B personality disorders.

As used interchangeably herein, a "potassium channel associated disorder" or a "54414 or 53763 associated disorder" include a disorder, disease or condition which is caused or characterized by a misregulation (*e.g.*, downregulation or upregulation) of 54414 or 53763 activity. 54414 or 53763 associated disorders can detrimentally affect cellular functions such as cellular proliferation, growth, differentiation, inter- or intra-cellular communication; tissue function, such as cardiac function or musculoskeletal function; systemic responses in an organism, such as nervous system responses, hormonal responses (*e.g.*, insulin response), or immune responses; and protection of cells from toxic compounds (*e.g.*, carcinogens, toxins, or mutagens).

In a preferred embodiment, 54414 or 53763 associated disorders include CNS disorders such as cognitive and neurodegenerative disorders, examples of which include, but are not limited to, Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, senile dementia, Huntington's disease, Gilles de la Tourette's syndrome, multiple sclerosis, amyotrophic lateral sclerosis, progressive *supranuclear* palsy, epilepsy, seizure disorders, and Jakob-Creutzfeldt disease; autonomic function disorders such as hypertension and sleep disorders, and neuropsychiatric disorders, such as depression, schizophrenia, schizoaffective disorder, korsakoff's psychosis, mania, anxiety disorders, or phobic disorders; learning or memory disorders, *e.g.*, amnesia or age-related memory loss, attention deficit disorder, dysthymic disorder, major depressive disorder, mania, obsessive-compulsive disorder, psychoactive substance use disorders, anxiety, phobias, panic disorder, as well as bipolar affective disorder, *e.g.*, severe bipolar affective (mood) disorder (BP-1), and bipolar affective neurological disorders, *e.g.*, migraine and obesity. Further CNS-related disorders include, for example, those listed in the American Psychiatric Association's Diagnostic and Statistical manual of Mental Disorders (DSM), the most current version of which is incorporated herein by reference in its entirety.

54414 or 53763 associated disorders also include cellular proliferation, growth, differentiation, or apoptosis disorders. Cellular proliferation, growth, differentiation, or apoptosis disorders include those disorders that affect cell proliferation, growth, differentiation, or apoptosis processes. As used herein, a "cellular proliferation, growth, differentiation, or apoptosis process" is a process by which a cell increases in number, size or content, by which a cell develops a specialized set of characteristics which differ from that of other cells, or by which a cell undergoes programmed cell death. The 54414 or 53763 molecules of the present invention may modulate cellular growth, proliferation, differentiation, or apoptosis, and may play a role in disorders characterized by aberrantly regulated growth, proliferation, differentiation, or apoptosis. Such disorders include cancer, *e.g.*, carcinoma, sarcoma, or leukemia; tumor angiogenesis and metastasis; skeletal dysplasia; hepatic disorders; and hematopoietic and/or myeloproliferative disorders.

Further examples of 54414 or 53763 associated disorders include cardiac-related disorders. Cardiovascular system disorders in which the 54414 or 53763 molecules of the invention may be directly or indirectly involved include arteriosclerosis, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism; tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, atrial fibrillation, Jervell syndrome, Lange-Nielsen syndrome, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, dilated cardiomyopathy, idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, and arrhythmia. 54414 or 53763 associated disorders also include disorders of the musculoskeletal system such as paralysis and muscle weakness, *e.g.*, ataxia, myotonia, and myokymia.

54414 or 53763 associated or related disorders also include hormonal disorders, such as conditions or diseases in which the production and/or regulation of hormones in an organism is aberrant. Examples of such disorders and diseases include type I and type II diabetes mellitus, pituitary disorders (*e.g.*, growth disorders), thyroid disorders (*e.g.*, hypothyroidism or hyperthyroidism), and reproductive or fertility disorders (*e.g.*, disorders which affect the organs of the reproductive system, *e.g.*, the prostate gland, the uterus, or the vagina; disorders which involve an imbalance in the levels of a reproductive hormone in a subject; disorders affecting the ability of a subject to reproduce; and disorders affecting secondary sex characteristic development, *e.g.*, adrenal hyperplasia).

54414 or 53763 associated or related disorders also include immune disorders, such as autoimmune disorders or immune deficiency disorders, *e.g.*, congenital X-linked infantile hypogammaglobulinemia, transient hypogammaglobulinemia, common variable immunodeficiency, selective IgA deficiency, chronic mucocutaneous candidiasis, or severe combined immunodeficiency.

As used interchangeably herein, a "phospholipid transporter associated disorder" or a "67076, 67102, 44181, 67084FL, or 67084alt associated disorder" includes a disorder, disease or condition which is caused or characterized by a misregulation (e.g., downregulation or upregulation) of 67076, 67102, 44181, 67084FL, or 67084alt activity. 67076, 67102, 44181, 67084FL, or 67084alt associated disorders can detrimentally affect cellular functions such as cellular proliferation, growth, differentiation, inter- or intra-cellular communication; tissue function, such as cardiac function or musculoskeletal function; systemic responses in an organism, such as nervous system responses, hormonal responses (e.g., insulin response), or immune responses; and protection of cells from toxic compounds (e.g., carcinogens, toxins, or mutagens). Examples of 67076, 67102, 44181, 67084FL, or 67084alt associated disorders include CNS disorders such as cognitive and neurodegenerative disorders, examples of which include, but are not limited to, Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, senile dementia, Huntington's disease, Gilles de la Tourette's syndrome, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, seizure disorders, and Jakob-Creutzfeldt disease; autonomic function disorders such as hypertension and sleep disorders, and neuropsychiatric disorders, such as depression, schizophrenia, schizoaffective disorder, korsakoff's psychosis, mania, anxiety disorders, or phobic disorders; learning or memory disorders, e.g., amnesia or age-related memory loss, attention deficit disorder, dysthymic disorder, major depressive disorder, mania, obsessive-compulsive disorder, psychoactive substance use disorders, anxiety, phobias, panic disorder, as well as bipolar affective disorder, e.g., severe bipolar affective (mood) disorder (BP-1), and bipolar affective neurological disorders, e.g., migraine and obesity. Further CNS-related disorders include, for example, those listed in the American Psychiatric Association's Diagnostic and Statistical manual of Mental Disorders (DSM), the most current version of which is incorporated herein by reference in its entirety.

Further examples of 67076, 67102, 44181, 67084FL, or 67084alt associated disorders include cardiac-related disorders. Cardiovascular system disorders in which the 67076, 67102, 44181, 67084FL, or 67084alt molecules of the invention may be directly or indirectly involved include arteriosclerosis, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, atrial fibrillation, Jervell syndrome, Lange-Nielsen syndrome, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, dilated cardiomyopathy, idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, and arrhythmia. 67076, 67102, 44181, 67084FL, or 67084alt associated disorders

also include disorders of the musculoskeletal system such as paralysis and muscle weakness, *e.g.*, ataxia, myotonia, and myokymia.

67076, 67102, 44181, 67084FL, or 67084alt associated disorders also include cellular proliferation, growth, or differentiation disorders. Cellular proliferation, growth, or differentiation disorders include those disorders that affect cell proliferation, growth, or differentiation processes. As used herein, a "cellular proliferation, growth, or differentiation process" is a process by which a cell increases in number, size or content, or by which a cell develops a specialized set of characteristics which differ from that of other cells. The 67076, 67102, 44181, 67084FL, or 67084alt molecules of the present invention are involved in phospholipid transport mechanisms, which are known to be involved in cellular growth, proliferation, and differentiation processes. Thus, the 67076, 67102, 44181, 67084FL, or 67084alt molecules may modulate cellular growth, proliferation, or differentiation, and may play a role in disorders characterized by aberrantly regulated growth, proliferation, or differentiation. Such disorders include cancer, *e.g.*, carcinoma, sarcoma, or leukemia; tumor angiogenesis and metastasis; skeletal dysplasia; hepatic disorders; and hematopoietic and/or myeloproliferative disorders.

67076, 67102, 44181, 67084FL, or 67084alt associated or related disorders also include hormonal disorders, such as conditions or diseases in which the production and/or regulation of hormones in an organism is aberrant. Examples of such disorders and diseases include type I and type II diabetes mellitus, pituitary disorders (*e.g.*, growth disorders), thyroid disorders (*e.g.*, hypothyroidism or hyperthyroidism), and reproductive or fertility disorders (*e.g.*, disorders which affect the organs of the reproductive system, *e.g.*, the prostate gland, the uterus, or the vagina; disorders which involve an imbalance in the levels of a reproductive hormone in a subject; disorders affecting the ability of a subject to reproduce; and disorders affecting secondary sex characteristic development, *e.g.*, adrenal hyperplasia).

67076, 67102, 44181, 67084FL, or 67084alt associated or related disorders also include immune disorders, such as autoimmune disorders or immune deficiency disorders, *e.g.*, congenital X-linked infantile hypogammaglobulinemia, transient hypogammaglobulinemia, common variable immunodeficiency, selective IgA deficiency, chronic mucocutaneous candidiasis, or severe combined immunodeficiency.

8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt associated or related disorders also include disorders affecting tissues in which 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt protein is expressed.

In addition, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides can be used to screen for naturally occurring 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrates, to screen for drugs or compounds which modulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or

67084alt activity, as well as to treat disorders characterized by insufficient or excessive production of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or production of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide forms which have decreased, aberrant or unwanted activity compared to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt wild type polypeptide (e.g., sugar transporter associated disorder, potassium channel associated disorders, a phospholipid transporter-associated disorders). Moreover, the anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibodies of the invention can be used to detect and isolate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, to regulate the bioavailability of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, and modulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides, have a stimulatory or inhibitory effect on, for example, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity is determined.

Determining the ability of the test compound to modulate 8099 or 46455 activity can be accomplished by monitoring, for example, intracellular or extracellular D-glucose, D-fructose, D-galactose, and/or mannose concentration, or insulin or glucagon secretion. The cell, for example, can be of mammalian origin, *e.g.*, a liver cell, fat cell, muscle cell, or a blood cell, such as an erythrocyte.

Determining the ability of the test compound to modulate 54414 or 53763 activity can be accomplished by monitoring, for example, potassium current, neurotransmitter release, and/or membrane excitability in a cell which expresses 54414 or 53763. The cell, for example, can be of mammalian origin, *e.g.*, a neuronal cell.

Determining the ability of the test compound to modulate 67076, 67102, 44181, 67084FL, or 67084alt activity can be accomplished by monitoring, for example, (i) interaction of 67076, 67102, 44181, 67084FL, or 67084alt with a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (*e.g.*, a phospholipid, ATP, or a non-67076, 67102, 44181, 67084FL, or 67084alt protein); (ii) transport of a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule (*e.g.*, an aminophospholipid such as phosphatidylserine or phosphatidylethanolamine) from one side of a cellular membrane to the other; (iii) the ability of 67076, 67102, 44181, 67084FL, or 67084alt to be phosphorylated or dephosphorylated; (iv) adoption by 67076, 67102, 44181, 67084FL, or 67084alt of an E1 conformation or an E2 conformation; (v) conversion of a 67076, 67102, 44181, 67084FL, or 67084alt substrate or target molecule to a product (*e.g.*, hydrolysis of

ATP); (vi) interaction of 67076, 67102, 44181, 67084FL, or 67084alt with a second non-67076, 67102, 44181, 67084FL, or 67084alt protein; (vii) modulation of substrate or target molecule location (e.g., modulation of phospholipid location within a cell and/or location with respect to a cellular membrane); (viii) maintenance of aminophospholipid gradients; 5 (ix) modulation of intra- or intercellular signaling and/or gene transcription (e.g., either directly or indirectly); and/or (x) modulation of cellular proliferation, growth, differentiation, apoptosis, absorption, and/or secretion.

The ability of the test compound to modulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt binding to a substrate or to bind to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can also be determined. 10 Determining the ability of the test compound to modulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt binding to a substrate can be accomplished, for example, by coupling the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate with a radioisotope or enzymatic label such that binding of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can be determined by 15 detecting the labeled 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate in a complex. Alternatively, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt could be coupled with a radioisotope or enzymatic label to 20 monitor the ability of a test compound to modulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt binding to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate in a complex. Determining the ability of the test compound to bind 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can be accomplished, for example, by coupling the 8099, 46455, 54414, 53763, 25 67076, 67102, 44181, 67084FL, or 67084alt substrate with a radioisotope or enzymatic label such that binding of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can be determined by detecting the labeled 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate in a complex. Alternatively, 8099, 46455, 30 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt binding to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate in a complex. Determining the ability of the test compound to bind 8099, 46455, 54414, 53763, 67076, 35 67102, 44181, 67084FL, or 67084alt can be accomplished, for example, by coupling the compound with a radioisotope or enzymatic label such that binding of the compound to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can be determined by detecting the labeled 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or

67084alt compound in a complex. For example, compounds (e.g., 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrates) can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a compound (e.g., a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate) to interact with 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt without the labeling of either the compound or the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule (e.g., a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate) with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule. Determining the ability of the test compound to modulate the activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule can be accomplished, for example, by determining the cellular location of the target molecule, or by determining whether the target molecule (e.g., ATP) has been hydrolyzed.

Determining the ability of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, or a biologically active fragment thereof, to bind to or interact with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide to bind to or interact with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting the cellular

location of target molecule, detecting catalytic/enzymatic activity of the target molecule upon an appropriate substrate, detecting induction of a metabolite of the target molecule (e.g., detecting the products of ATP hydrolysis, changes in intracellular K⁺ levels) detecting the induction of a reporter gene (comprising a target-responsive regulatory element
5 operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a target-regulated cellular response (i.e., membrane excitability, or cell growth, proliferation, differentiation, or apoptosis, sugar transport).

In yet another embodiment, an assay of the present invention is a cell-free assay in which a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt
10 polypeptide or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or biologically active portion thereof is determined. Preferred biologically active portions of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides to be used in assays of the
15 present invention include fragments which participate in interactions with non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt 3 molecules, e.g., fragments with high surface probability scores (see, for example, Figures 2, 9, 13, 17, 21, 25, 29, 33, and 37). Binding of the test compound to the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can be determined either directly or indirectly as
20 described above. In a preferred embodiment, the assay includes contacting the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or biologically active portion thereof with a known compound which binds 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact
25 with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, wherein determining the ability of the test compound to interact with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide comprises determining the ability of the test compound to preferentially bind to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt or biologically active portion thereof as
30 compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the 8099, 46455, 54414, 53763, 67076,
35 67102, 44181, 67084FL, or 67084alt polypeptide or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can be accomplished, for example, by determining the ability of the 8099, 46455, 54414, 53763,

67076, 67102, 44181, 67084FL, or 67084alt polypeptide to bind to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule by one of the methods described above for determining direct binding. Determining the ability of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide to bind to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can be accomplished by determining the ability of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide to further modulate the activity of a downstream effector of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule. For example, the activity of the effector molecule on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or biologically active portion thereof with a known compound which binds the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, wherein determining the ability of the test compound to interact with the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide comprises determining the ability of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide to preferentially bind to or modulate the activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of isolated proteins (e.g., 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt proteins or biologically active portions thereof). In the case of cell-free assays in which a membrane-bound form of an isolated protein is used it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the isolated protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-

dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, or interaction of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized micrometer plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or micrometer plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, substrate, or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with

8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or target molecules but which do not interfere with binding of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide to its target molecule can be derivatized to the wells of the plate, and unbound target or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or target molecule.

In another embodiment, modulators of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or polypeptide in the cell is determined. The level of expression of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or polypeptide in the presence of the candidate compound is compared to the level of expression of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or polypeptide in the absence of the candidate compound. The candidate compound can then be identified as a modulator of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression based on this comparison. For example, when expression of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or polypeptide is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or polypeptide expression. Alternatively, when expression of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or polypeptide is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or polypeptide expression. The level of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or polypeptide expression in the cells can be determined by methods described herein for detecting 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or polypeptide.

In yet another aspect of the invention, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993)

Biotechniques 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt ("8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -binding proteins" or "8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -bp") and are involved in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity. Such 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -binding proteins are also likely to be involved in the propagation of signals by the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptides or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt targets as, for example, downstream elements of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -mediated signaling pathway. Alternatively, such 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -binding proteins are likely to be 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for cellular transformation and/or tumorigenesis.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent

identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt modulating agent, an antisense 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecule, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -specific antibody, or a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences, described herein, can be used to map the location of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt genes on a chromosome. The mapping of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences. Computer analysis of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene

corresponding to the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes.

Other mapping strategies which can similarly be used to map a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for

marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

5 Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through
10 linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature*, 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, can be determined. If a mutation is observed in some or
15 all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of
20 genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt
25 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does
30 not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected
35 portions of an individual's genome. Thus, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 or SEQ ID NO:4 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 or SEQ ID NO:6 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated

fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 or SEQ ID NO:4 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique.

Examples of polynucleotide reagents include the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, or SEQ ID NO:13, having a length of at least 20 bases, preferably at least 30 bases.

The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide and/or nucleic acid expression as well as 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, nucleic acid expression or activity. For example, mutations in a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a

disorder characterized by or associated with 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid (e.g., mRNA, or genomic DNA) that encodes 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide such that the presence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid is detected in the biological sample. In another aspect, the present invention provides a method for detecting the presence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity such that the presence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity is detected in the biological sample. A preferred agent for detecting 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or genomic DNA. The nucleic acid probe can be, for example, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____, _____, _____, _____, or _____, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide is an antibody capable of binding to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA, polypeptide, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of PLTR polypeptide include introducing into a subject a labeled 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide; (ii) aberrant expression of a gene encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide; (iii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, wherein a wild-type form of the gene encodes a polypeptide with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity. "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes, but is not limited to, expression at non-wild type levels (e.g., over or under

expression); a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed (*e.g.*, increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage); a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene (*e.g.*, a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus).

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, mRNA, or genomic DNA, such that the presence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, mRNA or genomic DNA in the control sample with the presence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or mRNA in a biological sample; means for determining the amount of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt in the sample; and means for comparing the amount of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid.

2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt

expression or activity. As used herein, the term "aberrant" includes a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity which deviates from the wild type 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity. Aberrant expression or activity includes increased or decreased expression or activity, as well as expression or activity which does not follow the wild type developmental pattern of expression or the subcellular pattern of expression. For example, aberrant 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity is intended to include the cases in which a mutation in the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene causes the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene to be under-expressed or over-expressed and, situations in which such mutations result in a non-functional 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or a polypeptide which does not function in a wild-type fashion, e.g., a protein which does not interact with or transport a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate, or one which interacts with or transports a non-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as deregulated cellular proliferation. For example, the term unwanted includes a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity which is undesirable in a subject.

The assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with a misregulation in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide activity or nucleic acid expression, such as a cell growth, proliferation and/or differentiation disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disorder associated with a misregulation in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide activity or nucleic acid expression, such as a cell growth, proliferation and/or differentiation disorder, a sugar transporter associated disorder, or a potassium channel associated disorder, as described herein. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity in which a test sample is obtained from a subject and 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid (e.g., mRNA or genomic DNA) is detected, wherein the presence of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity.

As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a sugar transporter-associated disorder, a potassium channel associated disorder, or phospholipid transporter-associated disorder. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity in which a test sample is obtained and 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid expression or activity is detected (e.g., wherein the abundance of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity).

The methods of the invention can also be used to detect genetic alterations in a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide activity or nucleic acid expression, such as a cell growth, proliferation and/or differentiation disorder. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -polypeptide, or the mis-expression of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene; 2) an addition of one or more nucleotides to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene; 3) a substitution of one or more nucleotides of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene; 4) a chromosomal rearrangement of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene; 5) an alteration in the level of a messenger RNA transcript of a 8099, 46455, 54414, 53763, 67076, 67102,

44181, 67084FL, or 67084alt gene, 6) aberrant modification of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, 8) a non-wild type level of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -polypeptide, 9) allelic loss of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, and 10) inappropriate post-translational modification of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -polypeptide. As described herein, there are a large number of assays known in the art which can be used for detecting alterations in a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene (see Abravaya *et al.* (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene under conditions such that hybridization and amplification of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene and detect mutations by comparing the sequence of the sample 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene include methods in which protection from

cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence, *e.g.*, a wild-type 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence; the

resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene.

Furthermore, any cell type or tissue in which 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt is expressed may be utilized in the prognostic assays described herein.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide (*e.g.*, the modulation of gene expression, cellular signaling, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity, phospholipid transporter activity, and/or cell growth, proliferation, differentiation, absorption, and/or secretion mechanisms) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene expression, polypeptide levels, or upregulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity, can be monitored in clinical trials of subjects exhibiting decreased 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene expression, polypeptide levels, or downregulated 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene expression, polypeptide levels, or downregulate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity, can be monitored in clinical trials of subjects exhibiting increased 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene expression, polypeptide levels, or upregulated 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity. In such clinical trials, the expression or activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene, and preferably, other genes that have been implicated in, for example, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates 67076,

67102, 44181, 67084FL, or 67084alt activity (e.g., identified in a screening assay as described herein) can be identified.

Thus, to study the effect of agents on phospholipid transporter-associated disorders (e.g., disorders characterized by deregulated gene expression, cellular signaling, 67076, 5 67102, 44181, 67084FL, or 67084alt activity, phospholipid transporter activity, and/or cell growth, proliferation, differentiation, absorption, and/or secretion mechanisms), for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of 67076, 67102, 44181, 67084FL, or 67084alt and other genes implicated in the transporter-associated disorder, respectively. The levels of gene 10 expression (e.g., a gene expression pattern) can be quantified by northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of polypeptide produced, by one of the methods as described herein, or by measuring the levels of activity of 67076, 67102, 44181, 67084FL, or 67084alt or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells 15 to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate 20 identified by the screening assays described herein) including the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level 25 of expression or activity of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide, mRNA, or genomic DNA in the pre-administration sample with the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, 30 or 67084alt polypeptide, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, 35 decreased administration of the agent may be desirable to decrease expression or activity of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt to lower levels than detected, i.e. to decrease the effectiveness of the agent. According to such an embodiment, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt

expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

D. Methods of Treatment:

5 The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity, *e.g.* a phospholipid transporter-associated disorder.

"Treatment", as used herein, is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease or disorder, a symptom of disease or disorder or a predisposition toward a disease or disorder, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease or disorder, the symptoms of disease or disorder or the predisposition toward a disease or disorder. A therapeutic agent 10 includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application 20 of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug response phenotype", or "drug response genotype"). Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecules of the present invention or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid 30 treatment of patients who will experience toxic drug-related side effects.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity, by administering to the subject 35 a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt or an agent which modulates 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or at least one 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or

67084alt activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt aberrancy, for example, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt agonist or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell capable of expressing 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt with an agent that modulates one or more of the activities of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide activity associated with the cell, such that 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity in the cell is modulated. An agent that modulates 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide activity can be an agent as described herein, such as a nucleic acid or a polypeptide, a naturally-occurring target molecule of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide (e.g., a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt substrate), a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibody, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt agonist or antagonist, a peptidomimetic of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activities. Examples of such stimulatory agents include active 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide and a nucleic acid molecule encoding 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt that has been introduced into the cell. In another embodiment, the agent inhibits one or more 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activities. Examples of

such inhibitory agents include antisense 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt nucleic acid molecules, anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt 3 antibodies, and 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity. In another embodiment, the method involves administering a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression or activity.

Stimulation of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity is desirable in situations in which 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt is abnormally downregulated and/or in which increased 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity is likely to have a beneficial effect. Likewise, inhibition of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity is desirable in situations in which 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt is abnormally upregulated and/or in which decreased 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity is likely to have a beneficial effect.

3. Pharmacogenomics

The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity (e.g., 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically), for example, disorders characterized by aberrant 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene expression, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity, membrane excitability or conductance, gene transcription, phospholipid transporter activity, cellular signaling, and/or cell growth, proliferation, differentiation, absorption, and/or secretion disorders associated with aberrant or unwanted

8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecule or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecule or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11): 983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular

pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drugs target is known (*e.g.*, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecule or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecule or 8099, 46455, 54414,

53763, 67076, 67102, 44181, 67084FL, or 67084alt modulator, such as a modulator identified by one of the exemplary screening assays described herein.

4. Use of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt Molecules as Surrogate Markers

The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the

drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (e.g., a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt antibodies may be employed in an immune-based detection system for a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide marker, or 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -specific radiolabeled probes may be used to detect a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

The 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35(12): 1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or polypeptide (e.g., 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt DNA may correlate 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt drug response. The use of

pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

VI. Electronic Apparatus Readable Media and Arrays

5 Electronic apparatus readable media comprising 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information is also provided. As used herein, "8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information" refers to any nucleotide and/or amino acid sequence information particular to the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt molecules of the present invention, including but not limited to full-length nucleotide and/or amino acid sequences, partial nucleotide and/or amino acid sequences, polymorphic sequences including single nucleotide polymorphisms (SNPs), epitope sequences, and the like. Moreover, information "related to" said 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information includes detection of the presence or absence of a sequence (e.g., detection of expression of a sequence, fragment, polymorphism, etc.), determination of the level of a sequence (e.g., detection of a level of expression, for example, a quantitative detection), detection of a reactivity to a sequence (e.g., detection of protein expression and/or levels, for example, using a sequence-specific antibody), and the like. As used herein, "electronic apparatus readable media" refers to any suitable medium for storing, holding or containing data or information that can be read and accessed directly by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information of the present invention.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as a personal digital assistants (PDAs), cellular phone, pager and the like; and local and distributed processing systems.

As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate

manufactures comprising the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information.

A variety of software programs and formats can be used to store the sequence information on the electronic apparatus readable medium. For example, the sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like, as well as in other forms. Any number of dataprocessor structuring formats (e.g., text file or database) may be employed in order to obtain or create a medium having recorded thereon the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information.

By providing 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information in readable form, one can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the sequence information in readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder or a pre-disposition to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder, wherein the method comprises the steps of determining 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information associated with the subject and based on the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information, determining whether the subject has a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder or a pre-disposition to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder and/or recommending a particular treatment for the disease, disorder or pre-disease condition.

The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder or a pre-disposition to a disease associated with a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt wherein the method comprises the steps of determining 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information associated with the subject, and based on the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence information, determining whether the subject has a 8099, 46455, 54414, 53763,

67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder or a pre-disposition to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder, and/or recommending a particular treatment for the disease, disorder or pre-disease condition. The method may further comprise the step of receiving
5 phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

The present invention also provides in a network, a method for determining whether a subject has a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder or a pre-disposition to a 8099, 46455, 54414, 53763, 67076,
10 67102, 44181, 67084FL, or 67084alt -associated disease or disorder associated with 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt, said method comprising the steps of receiving 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or, 67084alt sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from
15 the network corresponding to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt and/or a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder, and based on one or more of the phenotypic information, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt information (e.g., sequence information and/or information related thereto), and the acquired
20 information, determining whether the subject has a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder or a pre-disposition to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

25 The present invention also provides a business method for determining whether a subject has a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder or a pre-disposition to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder, said method comprising the steps of receiving information related to 8099, 46455, 54414, 53763, 67076,
30 67102, 44181, 67084FL, or 67084alt (e.g., sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt and/or related to a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder, and based on one or more of the
35 phenotypic information, the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt information, and the acquired information, determining whether the subject has a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder or a pre-disposition to a 8099, 46455, 54414, 53763, 67076, 67102, 44181,

67084FL, or 67084alt -associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The invention also includes an array comprising a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

In addition to such qualitative determination, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder, progression of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder, and processes, such a cellular transformation associated with the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -associated disease or disorder.

The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (e.g., ascertaining the effect of 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt expression on the expression of other genes). This provides, for example, for a selection of alternate

molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including 5 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt) that could serve as a molecular target for diagnosis or therapeutic intervention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Figures and the Sequence Listing, are incorporated herein by reference.

EXAMPLES

10 **EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF HUMAN 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, AND 67084alt cDNAs**

In this example, the identification and characterization of the gene encoding human
15 8099, 46455, 54414, 53763, 67076, 67102, 44181, full length 67084 (67084FL), and
67084alt is described.

Isolation of the human 8099 and 46455 cDNAs

The invention is based, at least in part, on the discovery of a human gene encoding a
20 novel polypeptide, referred to herein as human 8099. The entire sequence of the human
clone 8099 was determined and found to contain an open reading frame termed human
"8099." The nucleotide sequence of the human 8099 gene is set forth in Figures 1A-B and
in the Sequence Listing as SEQ ID NO:1. The amino acid sequence of the human 8099
expression product is set forth in Figure 1 and in the Sequence Listing as SEQ ID NO: 2.
25 The 8099 polypeptide comprises 617 amino acids. The coding region (open reading frame)
of SEQ ID NO:1 is set forth as SEQ ID NO:3. Clone 8099, comprising the coding region of
human 8099, was deposited with the American Type Culture Collection (ATCC®), 10801
University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession
No. _____.

30 The invention is further based, at least in part, on the discovery of a human gene
encoding a novel polypeptide, referred to herein as human 46455. The entire sequence of
the human clone 46455 was determined and found to contain an open reading frame termed
human "46455." The nucleotide sequence of the human 46455 gene is set forth in Figure 4
and in the Sequence Listing as SEQ ID NO:4. The amino acid sequence of the human
35 46455 expression product is set forth in Figures 8A-B and in the Sequence Listing as SEQ
ID NO:5. The 46455 polypeptide comprises 528 amino acids. The coding region (open
reading frame) of SEQ ID NO:4 is set forth as SEQ ID NO:6. Clone 46455, comprising the
coding region of human 46455, was deposited with the American Type Culture Collection

(ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

Analysis of the Human 8099 and 46455 Molecules

5 A search using the polypeptide sequence of SEQ ID NO:2 was performed against the HMM database in PFAM (Figures 3A-B) resulting in the identification of a sugar transporter family domain in the amino acid sequence of human 8099 at about residues 43-564 of SEQ ID NO:2 (score = 318.2), a potential FecCD family domain in the amino acid sequence of human 8099 at about residues 26-227 of SEQ ID NO:2 (score = -218.2), and a
10 potential monocarboxylate transporter domain in the amino acid sequence of human 8099 at about residues 29-567 of SEQ ID NO:2 (score = -235.8).

The amino acid sequence of human 8099 was analyzed using the program PSORT (available through the Prosite website) to predict the localization of the proteins within the cell. This program assesses the presence of different targeting and localization amino acid
15 sequences within the query sequence. The results of this analysis show that human 8099 may be localized to the endoplasmic reticulum or mitochondria.

Searches of the amino acid sequence of human 8099 were further performed against the Prosite database. These searches resulted in the identification in the amino acid sequence of human 8099 of a number of potential N-glycosylation sites at about amino acid
20 residues 371-374, 383-386, 396-399, 401-404 of SEQ ID NO:2, a number of potential protein kinase C phosphorylation sites at about amino acid residues 220-222, 256-258, 403-405 of SEQ ID NO:2, a number of potential casein kinase II phosphorylation sites at about amino acid residues 18-21, 75-78, 169-172, 246-249, 256-259, 264-267, 385-388, 403-406, 443-446, 520-523 of SEQ ID NO:2, a number of potential N-myristoylation sites at about
25 amino acid residues 51-56, 59-64, 89-94, 141-146, 165-170, 178-183, 207-212, 228-233, 395-400, 441-446, and 493-498 of SEQ ID NO:2, a potential amidation site at about amino acid residues 104-107 of SEQ ID NO:2, a potential leucine zipper motif at about amino acid residues 112-133 of SEQ ID NO:2, and potential sugar transport protein signature 1 domain at about amino acid residues 97-114 of SEQ ID NO:2.

30 A MEMSAT analysis of the polypeptide sequence of SEQ ID NO:2 was also performed, predicting thirteen transmembrane domains in the amino acid sequence of human 8099 (SEQ ID NO:2) at about residues 32-49, 58-74, 81-101, 109-130, 138-156, 165-184, 198-217, 279-301, 315-338, 346-364, 463-487, 499-521, and 529-549. Further analysis of the amino acid sequence of SEQ ID NO:2 (e.g., alignment with, for example,
35 known *E. coli* sugar symporter proteins and a known human facilitative glucose transporter protein) showed that the second transmembrane domain at about amino acid residues 58-74 of SEQ ID NO:2 is not utilized, resulting in the presence of twelve transmembrane domains in the amino acid sequence of human 8099 (SEQ ID NO:2).

A search of the amino acid sequence of human 8099 was also performed against the ProDom database resulting in the identification of several transmembrane domains, a glycosyltransferase domain, and a sugar transport domain in the amino acid sequence of SEQ ID NO:2.

5 The human 8099 amino acid sequence was aligned with the amino acid sequence of the galactose-proton symporter GALP from *E. coli* using the CLUSTAL W (1.74) multiple sequence alignment program. The results of the alignment are set forth in Figure 4. The human 8099 amino acid sequence was also aligned with the amino acid sequence of the arabinose-proton symporter ARAE from *E. coli* using the CLUSTAL W (1.74) multiple
10 sequence alignment program. The results of the alignment are set forth in Figure 5. The human 8099 amino acid sequence was also aligned with the amino acid sequence of the facilitative glucose transporter GLUT8 from *Homo sapiens* using the CLUSTAL W (1.74) multiple sequence alignment program. The results of the alignment are set forth in Figure 7. Based on its homology to GLUT8, 8099 is also referred to herein as "GLUT8 homologue"
15 or "GLUT8h" and can be used interchangeably throughout.

A search using the polypeptide sequence of human 46455 (SEQ ID NO:5) was performed against the HMM database in PFAM (Figures 10A-C) resulting in the identification of a sugar transporter family domain in the amino acid sequence of human 46455 at about residues 58-469 of SEQ ID NO:5 (score = -63.4), a potential
20 sodium:galactoside symporter family domain in the amino acid sequence of human 46455 at about residues 212-505 of SEQ ID NO:5 (score = -121.2), and a potential monocarboxylate transporter domain in the amino acid sequence of human 46455 at about residues 60-473 of SEQ ID NO:5 (score = -208.2).

The amino acid sequence of human 46455 was analyzed using the program PSORT
25 to predict the localization of the proteins within the cell. This program assesses the presence of different targeting and localization amino acid sequences within the query sequence. The results of this analysis show that human 46455 may be localized to the endoplasmic reticulum, mitochondria, nucleus or secretory vesicles.

Searches of the amino acid sequence of human 46455 were further performed against
30 the Prosite database. These searches resulted in the identification in the amino acid sequence of human 46455 of a potential N-glycosylation site at about amino acid residues 199-202 of SEQ ID NO:5, a potential cAMP- and cGMP-dependent protein kinase C phosphorylation site at about amino acid residues 414-417 of SEQ ID NO:5, a number of potential protein kinase C phosphorylation sites at about amino acid residues 344-346, 413-
35 415, 442-444, and 518-520 of SEQ ID NO:5, a number of potential casein kinase II phosphorylation sites at about amino acid residues 11-14, 943-946, 959-962, 983-986, 1074-1077, 1108-1111, and 1112-1115 of SEQ ID NO:5, a number of potential N-myristoylation sites at about amino acid residues 89-94, 106-111, 288-293, 679-684, 767-772, 847-852, and

933-938 of SEQ ID NO:5, an amidation site at about amino acid residues 223-226 of SEQ ID NO:5, and an ATP/GTP-binding site motif A (P-loop) at about amino acid residues 1008-1015 of SEQ ID NO:5.

A MEMSAT analysis of the polypeptide sequence of SEQ ID NO:5 was also performed, predicting eleven transmembrane domains in the amino acid sequence of human 46455 (SEQ ID NO:5) at about residues 98-118, 126-145, 165-181, 188-205, 218-238, 273-294, 323-341, 357-377, 386-410, 423-441, and 462-485. Further analysis of the amino acid sequence of SEQ ID NO:5 (e.g., analysis of the hydropathy plot set forth in Figure 9) resulted in the identification of a twelfth transmembrane domain at about amino acid residues 58-74 of SEQ ID NO:5.

A search of the amino acid sequence of human 46455 was also performed against the ProDom database resulting in the identification of a transmembrane efflux domain in the amino acid sequence of SEQ ID NO:5.

The human 46455 amino acid sequence was aligned with the amino acid sequence of Z92825 from *C. elegans* using the CLUSTAL W (1.74) multiple sequence alignment program. The results of the alignment are set forth in Figure 11.

Isolation of the human 54414 and 53763 cDNA

The invention is based, at least in part, on the discovery of genes encoding novel members of the potassium channel family. The entire sequence of human clone Fbh54414 was determined and found to contain an open reading frame termed human "54414". The entire sequence of human clone Fbh53763 was determined and found to contain an open reading frame termed human "53763".

The nucleotide sequence encoding the human 54414 is shown in Figures 12A-D and is set forth as SEQ ID NO:7. The protein encoded by this nucleic acid comprises about 1118 amino acids and has the amino acid sequence shown in Figures 12A-D and set forth as SEQ ID NO:8. The coding region (open reading frame) of SEQ ID NO:7 is set forth as SEQ ID NO:9. Clone Fbh54414, comprising the coding region of human 54414, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

The nucleotide sequence encoding the human 53763 is shown in Figures 16A-C and is set forth as SEQ ID NO:10. The protein encoded by this nucleic acid comprises about 638 amino acids and has the amino acid sequence shown in Figures 16A-C and set forth as SEQ ID NO:11. The coding region (open reading frame) of SEQ ID NO:10 is set forth as SEQ ID NO:12. Clone Fbh53763, comprising the coding region of human 53763, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on ___, and assigned Accession No. _____.

Analysis of the human 54414 and 53763 Molecules

The amino acid sequences of human 54414 was analyzed using the program PSORT to predict the localization of the proteins within the cell. The results of the analyses show that human 54414 may be localized to the endoplasmic reticulum, the nucleus, secretory vesicles, or the mitochondria.

Analysis of the amino acid sequences of human 54414 was performed using MEMSAT. The amino acid sequence of human 54414 was also compared to the amino acid sequences of known potassium transporters (Figures 15A-B). This analysis resulted in the identification of six possible transmembrane domains in the amino acid sequence of human 54414 at residues 64-83, 104-127, 135-153, 161-173, 199-217, and 257-274 of SEQ ID NO:8 (Figure 13).

Searches of the amino acid sequences of human 54414 were performed against the HMM database (Figure 14). These searches resulted in the identification of an "ion transport protein domain", at about residues 104-277 of SEQ ID NO:8 (score= 62.4).

Searches of the amino acid sequence of human 54414 were further performed against the Prosite™ database. These searches resulted in the identification of several possible N-glycosylation sites at about amino acids residues 66-69, 99-102, 290-293, 545-548, 554-557, 573-576, 981-984, and 1106-1109 of SEQ ID NO:8, several possible cAMP- and cGMP-dependent protein kinase phosphorylation sites at about amino acids residues 625-628, 994-997, 1002-1005, and 1100-1103 of SEQ ID NO:8, several possible protein kinase C phosphorylation sites at about amino acid residues 43-45, 59-61, 68-70, 126-128, 158-160, 254-256, 298-300, 308-310, 354-356, 385-387, 464-466, 605-607, 903-905, 939-941, 947-949, 1005-1007, 1012-1014, 1030-1032, and 1099-1101 of SEQ ID NO:8, several possible casein kinase II phosphorylation sites at about amino acid residues 43-46, 115-118, 338-341, 386-389, 393-396, 485-488, 556-559, 651-654, 655-658, 663-666, 784-787, 837-840, 867-870, 907-910, 926-929, 943-946, 959-962, 983-986, 1074-1077, 1108-1111, and 1112-1115 of SEQ ID NO:8, several possible N-myristoylation sites at about amino acid residues 89-94, 106-111, 288-293, 679-684, 767-772, 847-852, and 933-938 of SEQ ID NO:8, a possible amidation site at about amino acid residues 223-226 of SEQ ID NO:8, and a possible ATP/GTP-binding site motif A (P-loop) at about amino acid residues 1008-1015 of SEQ ID NO:8.

The amino acid sequence of human 53763 was analyzed using the program PSORT to predict the localization of the proteins within the cell. The results of the analyses further show that human 53763 may be localized to the endoplasmic reticulum, the mitochondria, or the nucleus.

Analysis of the amino acid sequences of human 53763 was performed using MEMSAT. The amino acid sequence of human 53763 was also compared to the amino acid sequences of known potassium transporters (Figure 19). This analysis resulted in the

identification of six possible transmembrane domains in the amino acid sequence of human 53763 at residues 230-248, 287-303, 314-335, 346-368, 382-402, and 451-473 of SEQ ID NO:11 (Figure 17).

Searches of the amino acid sequence of human 53763 were also performed against the HMM database (Figures 18A-B). These searches resulted in the identification of a "NADH-ubiquinone/plastoquinone oxidoreductase domain" at about residues 317-467 of SEQ ID NO:11 (score = -81.7), an "ion transport protein domain" at about residues 281-472 of SEQ ID NO:11 (score = 116.9), and a "K⁺ channel tetramerisation domain" at about residues 8-156 of SEQ ID NO:11 (score = 156.7).

Searches of the amino acid sequence of human 53763 were also performed against the PrositeTM database. These searches resulted in the identification in the amino acid sequence of human 53763 a number of potential N-glycosylation sites at amino acid residues 84-84, 259-262, 266-269, 518-521, and 536-539 of SEQ ID NO:11, a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at amino acid residues 561-564 of SEQ ID NO:11, protein kinase C phosphorylation sites at amino acid residues 21-23, 25-27, 86-88, 120-122, 155-157, 205-207, 224-226, 336-338, 374-376, and 564-566 of SEQ ID NO:11, casein kinase II phosphorylation sites at amino acid residues 17-20, 49-52, 146-149, 283-286, 378-381, 414-417, 520-523, 541-544, 546-549, 553-556, 564-567, and 579-582 of SEQ ID NO:11, and N-myristoylation sites at amino acid residues 31-36, 76-81, 83-88, 89-94, 142-147, 176-181, 191-196, 199-204, 407-412, 450-455, 477-482, 590-595, and 606-611 of SEQ ID NO:11.

Searches of the amino acid sequences of human 54414 and human 53763 were also performed against the ProDom database. A potassium ionic calcium activated domain and two potassium ionic subunits were identified in the amino acid sequence of 54414 (SEQ ID NO:8). Several transmembrane domains and transport family domains were identified in the amino acid sequence of 53763 (SEQ ID NO:11).

The amino acid sequences of human 54414 and human 53763 were further analyzed for the presence of a "pore domain", also known as a "P-region domain". A pore domain was identified in the amino acid sequence of human 54414 at about residues 229-250 of SEQ ID NO:8. A pore domain was identified in the amino acid sequence of human 53763 at about residues 426-441 of SEQ ID NO:11.

The amino acid sequences of human 54414 and human 53763 were also analyzed for the presence of a "potassium channel signature sequence motif" (see Joiner, W. J. *et al.* (1998) *Nat. Neurosci.* 1:462-469 and references cited therein). A potassium channel signature sequence motif was identified in the amino acid sequence of human 54414 at about residues 239-246 of SEQ ID NO:8. A potassium channel signature sequence motif was identified in the amino acid sequence of human 53763 at about residues 436-441 of SEQ ID NO:11.

The amino acid sequence of human 53763 was also analyzed for the presence of a "voltage sensor motif". A voltage sensor motif was identified in the amino acid sequence of human 53763 at about residues 348-363 of SEQ ID NO:11. Positively charged amino acid residues in the voltage sensor motif were identified about residues 348, 351, 354, 357, 360, and 363 of SEQ ID NO:5.

Isolation of the human 67076, 67102, 44181, 67084FL, or 67084alt cDNAs

The invention is based, at least in part, on the discovery of a human gene encoding novel polypeptides, referred to herein as human 67076, 67102, 44181, 67084FL, and 67084alt. The entire sequence of the human clone 67076 was determined and found to contain an open reading frame termed human "67076." The nucleotide sequence of the human 67076 gene is set forth in Figures 20A-E and in the Sequence Listing as SEQ ID NO:13. The amino acid sequence of the human 67076 expression product is set forth in Figures 20A-E and in the Sequence Listing as SEQ ID NO:14. The 67076 polypeptide comprises 1129 amino acids. The coding region (open reading frame) of SEQ ID NO:13 is set forth as SEQ ID NO:15. Clone 67076, comprising the coding region of human 67076, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

The entire sequence of the human clone 67102 as determined and found to contain an open reading frame termed human "67102." The nucleotide sequence of the human gene is set forth in Figures 24A-E and in the Sequence Listing as SEQ ID NO:16. The amino acid sequence of the human 67102 expression product is set forth in Figures 24A-E and in the Sequence Listing as SEQ ID NO:17. The 67102 polypeptide comprises 1426 amino acids. The coding region (open reading frame) of SEQ ID NO:16 is set forth as SEQ ID NO:18. Clone 67102, comprising the coding region of human 67102, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

The entire sequence of the human clone 44181 was determined and found to contain an open reading frame termed human "44181." The nucleotide sequence of the human 44181 gene is set forth in Figures 28A-E and in the Sequence Listing as SEQ ID NO:19. The amino acid sequence of the human 44181 expression product is set forth in Figure 7A-E and in the Sequence Listing as SEQ ID NO:20. The 44181 polypeptide comprises 1177 amino acids. The coding region (open reading frame) of SEQ ID NO:19 is set forth as SEQ ID NO:21. Clone 44181, comprising the coding region of human 44181, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

The entire sequence of the human clone 67084FL was determined and found to contain an open reading frame termed human "67084FL." The nucleotide sequence of the

human 67084FL gene is set forth in Figures 32A-D and in the Sequence Listing as SEQ ID NO:22. The amino acid sequence of the human 67084FL expression product is set forth in Figures 32A-D and in the Sequence Listing as SEQ ID NO:23. The 67084FL polypeptide comprises 1084 amino acids. The coding region (open reading frame) of SEQ ID NO:22 is set forth as SEQ ID NO:24. Clone 67084FL, comprising the coding region of human 67084FL, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

The entire sequence of the human clone 67084alt was determined and found to contain an open reading frame termed human "67084alt." The nucleotide sequence of the human 67084alt gene is set forth in Figures 36A-D and in the Sequence Listing as SEQ ID NO:25. The amino acid sequence of the human 67084alt expression product is set forth in Figures 36A-D and in the Sequence Listing as SEQ ID NO:26. The 67084alt polypeptide comprises 1095 amino acids. The coding region (open reading frame) of SEQ ID NO:25 is set forth as SEQ ID NO:27. Clone 67084alt, comprising the coding region of human 67084alt, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

Analysis of the Human 67076, 67102, 44181, 67084FL, or 67084alt Molecules

The amino acid sequences of human 67076, 67102, 44181, 67084FL, or 67084alt were analyzed for the presence of sequence motifs specific for P-type ATPases (as defined in Tang, X. et al. (1996) *Science* 272:1495-1497 and Fagan, M. J. and Saier, M. H. (1994) *J. Mol. Evol.* 38:57).

These analyses resulted in the identification of a P-type ATPase sequence I motif in the amino acid sequence of human 67076 at residues 173-181 of SEQ ID NO:14, in the amino acid sequence of human 67102 at residues 208-216 of SEQ ID NO:17, in the amino acid sequence of human 44181 at residues 173-181 of SEQ ID NO:20, in the amino acid sequence of human 67084FL at residues 213-221 of SEQ ID NO:23, and in the amino acid sequence of human 67084alt at residues 213-221 of SEQ ID NO:26.

These analyses also resulted in the identification of a P-type ATPase sequence 2 motif in the amino acid sequence of human 67076 at residues 406-415 of SEQ ID NO:14, in the amino acid sequence of human 67102 at residues 435-444 of SEQ ID NO:17, in the amino acid sequence of human 44181 at residues 404-413 of SEQ ID NO:20, in the amino acid sequence of human 67084FL at residues 413-422 of SEQ ID NO:23, and in the amino acid sequence of human 67084alt at residues 413-422 of SEQ ID NO:26.

These analyses further resulted in the identification of a P-type ATPase sequence 3 motif in the amino acid sequence of human 67076 at residues 813-824 of SEQ ID NO:14, in

the amino acid sequence of human 67102 at residues 1054-1064 of SEQ ID NO:17, in the amino acid sequence of human 44181 at residues 819-829 of SEQ ID NO:20, in the amino acid sequence of human 67084FL at residues 820-830 of SEQ ID NO:23, and in the amino acid sequence of human 67084alt at residues 820-830 of SEQ ID NO:26.

5 The amino acid sequences of human 67076, 67102, 44181, 67084FL, and 67084alt were also analyzed for the presence of phospholipid transporter specific amino acid residues (as defined in Tang, X. et al. (1996) *Science* 272:1495-1497). These analyses also resulted in the identification of phospholipid transporter specific amino acid residues in the amino acid sequence of human 67076 at residues 174, 177, 407, 813, 823, and 824 of SEQ ID
10 NO:14. These analyses resulted in the identification of phospholipid transporter specific amino acid residues 208, 209, 212, 436, 1054, 1063, and 1064 in the amino acid sequence of human 67102 at residues of SEQ ID NO:17. These analyses further resulted in the identification of phospholipid transporter specific amino acid residues 174, 177, 405, 819, 928, and 929 in the amino acid sequence of human 44181 at residues of SEQ ID NO:20.
15 These analyses further resulted in the identification of phospholipid transporter specific amino acid residues 214, 217, 820, 829, and 830 in the amino acid sequence of human 67084FL at residues of SEQ ID NO:23. These analyses still further resulted in the identification of phospholipid transporter specific amino acid residues 214, 217, 820, 829, and 830 in the amino acid sequence of human 67084alt at residues of SEQ ID NO:26.

20 The amino acid sequences of human 67076, 67102, 44181, 67084FL, and 67084alt were also analyzed for the presence of extramembrane domains. An N-terminal large extramembrane domain was identified in the amino acid sequence of human 67076 at residues 105-291 of SEQ ID NO:14. A C-terminal large extramembrane domain was identified in the amino acid sequence of human 67076 at residues 366-872 of SEQ ID
25 NO:14. An N-terminal large extramembrane domain was identified in the amino acid sequence of human 67102 at residues 141-321 of SEQ ID NO:17. A C-terminal large extramembrane domain was identified in the amino acid sequence of human 67102 at residues 391-581 of SEQ ID NO:17. An N-terminal large extramembrane domain was identified in the amino acid sequence of human 44181 at residues 105-289 of SEQ ID
30 NO:20. A C-terminal large extramembrane domain was identified in the amino acid sequence of human 44181 at residues 364-877 of SEQ ID NO:20. An N-terminal large extramembrane domain was identified in the amino acid sequence of human 67084FL at residues 145-330 of SEQ ID NO:23. A C-terminal large extramembrane domain was identified in the amino acid sequence of human 67084FL at residues 380-886 of SEQ ID
35 NO:23. An N-terminal large extramembrane domain was identified in the amino acid sequence of human 67084alt at residues 145-330 of SEQ ID NO:26. A C-terminal large extramembrane domain was identified in the amino acid sequence of human 67084alt at residues 380-886 of SEQ ID NO:26.

The amino acid sequence of human 67076 was analyzed using the program PSORT to predict the localization of the proteins within the cell. This program assesses the presence of different targeting and localization amino acid sequences within the query sequence. The results of this analysis predict that human 67076 may be localized to the endoplasmic reticulum.

Searches of the amino acid sequence of human 67076 were further performed against the Prosite database. These searches resulted in the identification in the amino acid sequence of human 67076 of a number of potential N-glycosylation sites at amino acid residues 121-124, 392-395, 761-764, 992-995, and 1098-1101 of SEQ ID NO:14, a number of potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at amino acid residues 135-138, 545-548, 1091-1094, and 1102-1105 of SEQ ID NO:14, a number of potential protein kinase C phosphorylation sites at amino acid residues 47-49, 138-140, 204-206, 250-252, 254-256, 278-280, 308-310, 328-330, 334-336, 408-410, 680-682, 701-703, 708-710, 733-735, 736-738, 1008-1010, 1094-1096, 1100-1102, 1109-1111, and 1113-1115 of SEQ ID NO:14, a number of casein kinase II phosphorylation sites at amino acid residues 30-33, 264-267, 282-285, 328-331, 413-416, 442-445, 449-452, 494-497, 646-649, 693-696, 704-707, 762-765, 813-816, 924-927, 982-985, and 1121-1124 of SEQ ID NO:14, a number of potential tyrosine kinase phosphorylation sites at amino acid residues 252-258, 739-747 of SEQ ID NO:14, a number of N-myristoylation sites at amino acid residues 388-393, 440-445, 482-487, 514-519, 564-569, 753-758, and 807-812 of SEQ ID NO:14, an ATP/GTP-binding site motif (P-loop) at amino acid residues 271-278 of SEQ ID NO:14, and an E1-E2 ATPases phosphorylation site at amino acid residues 409-415 of SEQ ID NO:14.

A MEMSAT analysis of the polypeptide sequence of SEQ ID NO:14 was also performed, predicting three potential transmembrane domains in the amino acid sequence of human 67076 (SEQ ID NO:14). However, a structural, hydrophobicity, and antigenicity analysis (Figure 21) resulted in the identification of ten transmembrane domains. Accordingly, the 67076 protein of SEQ ID NO:14 is predicted to have at least ten transmembrane domains, identified as transmembrane (TM) domains 1 through 10, at about residues 57-77, 84-105, 292-313, 345-365, 863-883, 905-926, 956-977, 989-1009, 1021-1041, and 1060-1087.

A search using the polypeptide sequence of SEQ ID NO:14 was performed against the HMM database in PFAM resulting in the identification of a potential hydrolase domain in the amino acid sequence of human 67076 at about residues 403-837 of SEQ ID NO:14 (score = 12.7).

A search of the amino acid sequence of human 67076 was also performed against the ProDom database resulting in the identification of several hydrolase domains and phosphorylation domains in the amino acid sequence of 67076 (SEQ ID NO:14).

The amino acid sequence of human 67102 was analyzed using the program PSORT. The results of this analysis predict that human 67102 may be localized to the endoplasmic reticulum.

5 Searches of the amino acid sequence of human 67102 were further performed against the Prosite database. These searches resulted in the identification in the amino acid sequence of human 67102 of a number of potential N-glycosylation sites at amino acid residues 29-32, 303-306, 1365-1368, and 1397-1400 of SEQ ID NO:17, a glycosaminoglycan attachment site at amino acid residues 526-529 of SEQ ID NO:17, a
10 number of potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at amino acid residues 38-41, 451-545, 635-638, and 777-780 of SEQ ID NO:17, a number of protein kinase C phosphorylation sites at amino acid residues 47-49, 78-80, 161-163, 240-242, 262-264, 280-282, 437-439, 500-502, 563-565, 633-635, 644-646, 695-697, 743-745, 774-776, 827-829, 1000-1002, 1360-1362, and 1371-1373 of SEQ ID NO:17, a number of
15 potential casein kinase II phosphorylation sites at amino acid residues 20-23, 161-164, 176-179, 184-187, 199-202, 210-213, 232-235, 241-244, 262-265, 312-315, 345-348, 405-408, 442-445, 471-474, 477-480, 543-546, 621-624, 644-647, 670-673, 693-696, 727-730, 850-853, 866-869, 892-895, 977-980, 1074-1077, 1141-1144, 1199-1202, 1221-1224, 1339-1342, 1399-1402, and 1403-1406 of SEQ ID NO:17, two tyrosine kinase phosphorylation
20 sites at amino acid residues 21-28 and 847-854 of SEQ ID NO:17, a number of potential N-myristoylation sites at amino acid residues 69-74, 341-346, 488-493, 510-515, 519-524, 525-530, 651-656, 703-708, 714-719, 901-906, 955-960, 992-997, 1070-1075, 1139-1144, 1229-1234, and 1261-1266 of SEQ ID NO:17, two potential amidation sites at amino acid residues 36-39 and 1371-1374 of SEQ ID NO:17, two ATP/GTP-binding site motif A (P-loop) at amino acid residues 996-1003 and 1364-1371, an E1-E2 ATPases phosphorylation
25 site at amino acid residues 438-444 of SEQ ID NO:17, and a prokaryotic membrane lipoprotein lipid attachment site at amino acid residues 26-36 of SEQ ID NO:17.

A MEMSAT analysis of the polypeptide sequence of SEQ ID NO:17 was also performed, predicting ten potential transmembrane domains in the amino acid sequence of
30 human 67102 (SEQ ID NO:17) at about residues 98-115, 122-140, 322-344, 366-390, 582-601, 752-770, 1145-1166, 1225-1246, 1253-1276, and 1298-1317.

A search using the polypeptide sequence of SEQ ID NO:17 was performed against the HMM database in PFAM resulting in the identification of a potential hydrolase domain in the amino acid sequence of human 67102 at about residues 432-1077 of SEQ ID NO:17
35 (score = 1.5), and the identification of a potential DUF6 domain in the amino acid sequence of human 67102 at about residues 1127-1271 of SEQ ID NO:17 (score = -24.6).

A search of the amino acid sequence of human 67102 was also performed against the ProDom database resulting in the identification of several hydrolase domains and phosphorylation domains in the amino acid sequence of 667102 (SEQ ID NO:17).

5 The amino acid sequence of human 44181 was analyzed using the program PSORT. The results of this analysis predict that human 44181 may be localized to the endoplasmic reticulum.

Searches of the amino acid sequence of human 44181 were further performed against the Prosite database. These searches resulted in the identification in the amino acid
10 sequence of human 44181 of a number of potential N-glycosylation sites at amino acid residues 331-334, 390-393, 449-452, 461-464, 477-480, 786-789, and 998-1001 of SEQ ID NO:20, a number of potential cAMP- and cGMP-dependent protein kinase phosphorylation sites at amino acid residues 577-580, 633-636, and 750-753 of SEQ ID NO:20, a number of protein kinase C phosphorylation sites at amino acid residues 46-48, 163-165, 276-278, 332-
15 334, 406-408, 470-472, 574-576, 636-638, 957-959, 1014-1016, and 1102-1104 of SEQ ID NO:20, a number of potential casein kinase C phosphorylation sites at amino acid residues 115-118, 262-265, 280-283, 411-414, 473-476, 520-523, 527-530, 636-639, 678-681, 737-740, 906-909, 929-932, 1100-1103, 1154-1157, and 1170-1173 of SEQ ID NO:20, a potential tyrosine kinase phosphorylation site at amino acid residues 316-322 of SEQ ID
20 NO:20, a number of potential N-myristoylation sites at amino acid residues 131-136, 596-601, 766-771, and 993-998 of SEQ ID NO:20, and an E1-E2 ATPases phosphorylation site at amino acid residues 407-413 of SEQ ID NO:20.

A MEMSAT analysis of the polypeptide sequence of SEQ ID NO:20 was also performed, predicting three potential transmembrane domains in the amino acid sequence of
25 human 44181 (SEQ ID NO:20). However, a structural, hydrophobicity, and antigenicity analysis (Figure 29) resulted in the identification of ten transmembrane domains. Accordingly, the 44181 protein (SEQ ID NO:20) is predicted to have at least ten transmembrane domains, which are identified as transmembrane (TM) domains 1 through 10, at about residues 56-72, 87-103, 290-311, 343-363, 878-898, 911-931, 961-982, 995-
30 1015, 1027-1047, and 1062-1086.

A search using the polypeptide sequence of SEQ ID NO:20 was performed against the HMM database in PFAM resulting in the identification of a potential E1-E2 ATPase domain in the amino acid sequence of human 44181 at about residues 126-164 of SEQ ID NO:20 (score = 8.6), the identification of a potential DUF132 domain in the amino acid
35 sequence of human 44181 at about residues 579-719 of SEQ ID NO:20 (score = -72.9), and the identification of a potential hydrolase domain in the amino acid sequence of human 44181 at about residues 401-842 of SEQ ID NO:20 (score = 42.8).

A search of the amino acid sequence of human 44181 was also performed against the ProDom database. A search of the amino acid sequence of human 44181 was also performed against the ProDom database resulting in the identification of several hydrolase domains and phosphorylation domains in the amino acid sequence of 44181 (SEQ ID NO:20).

5 A Clustal W (1.74) alignment of the amino acid sequence of human 44181 (SEQ ID NO:20) and human potential phospholipid-transporting ATPase IR (ATIR; GenBank Accession No.:Q9Y2G3) revealed some sequence homology between 44181 and Accession No.:Q9Y2G3.

10 The amino acid sequence of human 67084FL was analyzed using the program PSORT. The results of this analysis predict that human 67084FL may be localized to the endoplasmic reticulum.

Searches of the amino acid sequence of human 67084FL were further performed against the Prosite database. These searches resulted in the identification in the amino acid
15 sequence of human 67084FL of a number of potential N-glycosylation sites at amino acid residues 310-313, 464-467, and 529-532 of SEQ ID NO:23, a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at amino acid residues 1071-1074 of SEQ ID NO:23, a number of protein kinase C phosphorylation sites 82-84, 168-170, 204-206, 301-303, 371-373, 415-417, 486-488, 585-587, 666-668, 744-746, 800-802, 813-815, 872-874,
20 957-959, and 1009-1011 of SEQ ID NO:23, a number of potential casein kinase II phosphorylation sites at amino acid residues 265-268, 301-304, 402-405, 422-425, 535-538, 596-599, 661-664, 686-689, and 745-748 of SEQ ID NO:23, a tyrosine kinase phosphorylation site at amino acid residues 813-816 of SEQ ID NO:23, a number of potential N-myristoylation sites at amino acid residues 292-297, 462-467, 568-573, 606-611, 824-829, 887-892, and 1026-1031 of SEQ ID NO:23, a potential amidation site at
25 amino acid residues 813-816 of SEQ ID NO:23, a prokaryotic membrane lipoprotein lipid attachment site at amino acid residues 105-115, a leucine zipper pattern at amino acid residues 325-346, and an E1-E2 ATPases phosphorylation site at amino acid residues 416-422 of SEQ ID NO:23.

30 A MEMSAT analysis of the polypeptide sequence of SEQ ID NO:23 was also performed, predicting nine potential transmembrane domains in the amino acid sequence of human 67084FL (SEQ ID NO:23). However, a structural, hydrophobicity, and antigenicity analysis (Figure 33) resulted in the identification of ten transmembrane domains. Accordingly, the 67084FL protein of SEQ ID NO:23 is predicted to have at least ten
35 transmembrane domains, which are identified as transmembrane (TM) domains 1 through 10, at about residues 104-120, 124-144, 331-350, 357-374, 887-903, 912-931, 961-983, 990-1008, 1015-1035, and 1043-1067.

A search using the polypeptide sequence of SEQ ID NO:23 was performed against the HMM database in PFAM resulting in the identification of two potential E1-E2 ATPase in the amino acid sequence of human 67084FL at about residues 171-199 of SEQ ID NO:23 (score = 3.0) and 277-305 of SEQ ID NO:23 (score = 13.0), and a hydrolase domain at about residues 410-843 of SEQ ID NO:23 (score = 19.2).

A search of the amino acid sequence of human 67084FL was also performed against the ProDom database resulting in the identification of several hydrolase domains, phosphorylation domains, and ATPase domains in the amino acid sequence of 67084FL (SEQ ID NO:23).

A Clustal W (1.74) alignment of the amino acid sequence of human 67084FL (SEQ ID NO:23) and human membrane transport protein (MTRP-1; GenBank Accession No.: Y71056, International Publication No. WO 2000/26245-A2) revealed some sequence homology between 67084FL and Accession No.: Y71056.

The amino acid sequence of human 67084alt was analyzed using the program PSORT. The results of this analysis predict that human 67084alt may be localized to the endoplasmic reticulum.

Searches of the amino acid sequence of human 67084alt were further performed against the Prosite database. These searches resulted in the identification in the amino acid sequence of human 67084alt of a number of potential N-glycosylation sites at amino acid residues 310-313, 464-467, and 529-532 of SEQ ID NO:26, a potential cAMP- and cGMP-dependent protein kinase phosphorylation site at amino acid residues 1083-1086, a number of protein kinase C phosphorylation sites at amino acid residues 82-84, 168-170, 204-2-6, 301-303, 371-373, 415-417, 486-488, 585-587, 666-668, 744-746, 800-802, 813-815, 872-874, 957-959, and 1009-1011 of SEQ ID NO:26, a number of potential casein kinase II phosphorylation sites at amino acid residues 265-268, 301-304, 402-405, 422-445, 535-538, 596-599, 661-664, 686-689, and 745-748 of SEQ ID NO:26, a tyrosine kinase phosphorylation site at amino acid residues 641-648, a number of potential N-myristoylation sites at amino acid residues 292-297, 462-467, 568-573, 606-611, 824-829, 887-892, and 1026-1031 of SEQ ID NO:26, a potential amidation site at amino acid residues 813-816 of SEQ ID NO:26, a potential prokaryotic membrane lipoprotein lipid attachment site at amino acid residues 105-115 of SEQ ID NO:26, a leucine zipper pattern at amino acid residues 325-346 of SEQ ID NO:26, and an E1-E2 ATPases phosphorylation site at amino acid residues 416-422 of SEQ ID NO:26.

A MEMSAT analysis of the polypeptide sequence of SEQ ID NO:26 was also performed, predicting nine potential transmembrane domains in the amino acid sequence of human 67084alt (SEQ ID NO:26). However, a structural, hydrophobicity, and antigenicity analysis (Figure 37) resulted in the identification of ten transmembrane domains.

Accordingly, the 67084alt protein of SEQ ID NO:26 is predicted to have at least ten transmembrane domains, which are identified as transmembrane (TM) domains 1 through 10, at about residues 104-120, 124-144, 331-350, 357-374, 887-903, 912-931, 961-983, 990-1008, 1015-1035, and 1043-1067.

5 A search using the polypeptide sequence of SEQ ID NO:26 was performed against the HMM database in PFAM resulting in the identification of two potential E1-E2 ATPase in the amino acid sequence of human 67084alt at about residues 42-70 of SEQ ID NO:26 (score = 3.0) and 105-133 of SEQ ID NO:26 (score = 13.0), and a potential hydrolase domain at about amino acid residues 410-843 of SEQ ID NO:26 (score = 19.2).

10 A search of the amino acid sequence of human 67084alt was also performed against the ProDom database resulting in the identification of several hydrolase domains, phosphorylation domains, and ATPase domains in the amino acid sequence of 67084alt (SEQ ID NO:26).

A Clustal W (1.74) alignment of the amino acid sequence of human 67084alt (SEQ ID NO:14) and human membrane transport protein (MTRP-1; GenBank Accession No.:Y71056, International Publication No. WO 2000/26245-A2) revealed some sequence homology between 67084alt and Accession No.: Y71056.

EXAMPLE 2: TISSUE EXPRESSION OF HUMAN 8099, 46455, 54414, 53763, 67076, 67102, 44181; full length 67084 (67084FL), and 67084alt mRNA

Tissue Distribution of Human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, and 67084alt mRNA Using Taqman™ Analysis

This example describes the tissue distribution of human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt mRNA in a variety of cells and tissues, as determined using the TaqMan™ procedure. The Taqman™ procedure is a quantitative, reverse transcription PCR-based approach for detecting mRNA. The RT-PCR reaction exploits the 5' nuclease activity of AmpliTaq Gold™ DNA Polymerase to cleave a TaqMan™ probe during PCR. Briefly, cDNA was generated from the samples of interest, e.g., lung, ovary, colon, and breast normal and tumor samples, and used as the starting material for PCR amplification. In addition to the 5' and 3' gene-specific primers, a gene-specific oligonucleotide probe (complementary to the region being amplified) was included in the reaction (*i.e.*, the Taqman™ probe). The TaqMan™ probe includes the oligonucleotide with a fluorescent reporter dye covalently linked to the 5' end of the probe (such as FAM (6-carboxyfluorescein), TET (6-carboxy-4,7,2',7'-tetrachlorofluorescein), JOE (6-carboxy-4,5-dichloro-2,7-dimethoxyfluorescein), or VIC) and a quencher dye (TAMRA (6-carboxy-N,N,N',N'-tetramethylrhodamine) at the 3' end of the probe.

During the PCR reaction, cleavage of the probe separates the reporter dye and the

quencher dye, resulting in increased fluorescence of the reporter. Accumulation of PCR products is detected directly by monitoring the increase in fluorescence of the reporter dye. When the probe is intact, the proximity of the reporter dye to the quencher dye results in suppression of the reporter fluorescence. During PCR, if the target of interest is present, the probe specifically anneals between the forward and reverse primer sites. The 5'-3' nucleolytic activity of the AmpliTaq™ Gold DNA Polymerase cleaves the probe between the reporter and the quencher only if the probe hybridizes to the target. The probe fragments are then displaced from the target, and polymerization of the strand continues. The 3' end of the probe is blocked to prevent extension of the probe during PCR. This process occurs in every cycle and does not interfere with the exponential accumulation of product. RNA was prepared using the trizol method and treated with DNase to remove contaminating genomic DNA. cDNA was synthesized using standard techniques. Mock cDNA synthesis in the absence of reverse transcriptase resulted in samples with no detectable PCR amplification of the control gene confirms efficient removal of genomic DNA contamination.

Tissue Distribution of Human 8099

A human tissue panel was tested revealing highest expression of human 8099 mRNA in congestive heart failure (CHF) heart, normal prostate, and brain (see Table 1, below).

TABLE 1.

Tissue Type	Mean	± 2 Mean	ΔΔ Ct	Expression
Artery normal	30.83	22.31	8.52	2.7241
Aorta diseased	32.77	22.32	10.45	0.7149
Vein normal	29.41	20.23	9.18	1.724
Coronary SMC	31.2	20.91	10.3	0.7932
HUVEC	32.16	21.38	10.78	0.5687
Hemangioma	32.86	19.66	13.21	0.1059
Heart normal	28.05	20.43	7.62	5.0834
Heart CHF	26.98	20.68	6.3	12.6914
Kidney	27.76	20.45	7.3	6.3238
Skeletal Muscle	29.7	22.17	7.53	5.4294
Adipose normal	34.16	20.59	13.56	0.0828
Pancreas	33.23	22.29	10.94	0.5108
primary osteoblasts	32	20.61	11.39	0.3726
Osteoclasts (diff)	30.9	17.55	13.35	0.0958
Skin normal	34.12	22.45	11.68	0.3058
Spinal cord normal	31.93	21.07	10.87	0.5362
Brain Cortex normal	28.4	22.34	6.06	14.9885
Brain Hypothalamus normal	29.68	22.35	7.34	6.1936
Nerve	32.96	22.25	10.72	0.5949
DRG (Dorsal Root Ganglion)	30.81	22.15	8.65	2.4808
Breast normal	31.91	21.14	10.77	0.5747
Breast tumor	32.73	20.93	11.81	0.2785
Ovary normal	30.41	19.82	10.6	0.6465

Ovary Tumor	28.36	19.06	9.31	1.5755
Prostate Normal	27.29	19.77	7.52	5.4482

Tissue Distribution of Human 46455

- 5 A human vessel and tissue panel was tested revealing highest expression of human 46455 mRNA in human umbilical vein endothelial cells (HUVEC), erythroid cells, normal artery, megakaryocytes, kidney, and CHF heart. 46455 was expressed at higher levels in lung tumor, breast tumor, and colon tumor versus normal lung, breast and colon tissues, indicating a possible role for 46455 in cellular proliferation disorders (see Table 2, below).

10 **TABLE 2.**

Tissue Type	Mean	± 2 Mean	ΔΔ Ct	Expression
Artery normal	28.05	24.09	2.67	157.6722
Aorta diseased	28.75	23.66	3.79	72.0429
Vein normal	27.75	21.72	4.75	37.1627
Coronary SMC	28.2	25.12	1.78	290.176
HUVEC	24.18	22.59	0.29	817.9021
Hemangioma	25.15	20.98	2.88	135.8419
Heart normal	26.44	21.82	3.33	99.4421
Heart CHF	25.54	21.09	3.15	112.6563
Kidney	25.98	21.49	3.2	108.8188
Skeletal Muscle	28.22	24.14	2.79	144.586
Adipose normal	28.38	22.21	4.88	33.9605
Pancreas	27.91	23.22	3.4	94.4045
primary osteoblasts	27.11	21.85	3.97	63.8133
Osteoclasts (diff)	23.64	18.8	3.55	85.3775
Skin normal	28.43	23.27	3.88	68.1567
Spinal cord normal	26.88	22.12	3.47	90.2456
Brain Cortex normal	26.42	23.4	1.73	301.452
Brain Hypothalamus normal	28.1	23.55	3.26	104.386
Nerve	28.59	23.88	3.43	92.7827
DRG (Dorsal Root Ganglion)	28.33	23.76	3.28	102.9489
Breast normal	27.31	22.32	3.7	76.9465
Breast tumor	26.47	22.11	3.07	119.0797
Ovary normal	26.59	22.16	3.13	113.8337
Ovary Tumor	28.47	21.84	5.33	24.8605
Prostate Normal	27.09	21.68	4.12	57.5117
Prostate Tumor	26.51	21.58	3.64	80.2141
Salivary glands	27.16	20.81	5.07	29.8733
Colon normal	26.3	20	5	31.1419
Colon Tumor	25.09	20.52	3.29	102.5927
Lung normal	26.02	19.75	4.98	31.6862
Lung tumor	25.09	21.31	2.48	178.6243
Lung COPD	25.26	19.71	4.26	52.193
Colon IBD	26.3	18.91	6.1	14.5786
Liver normal	27.66	21.8	4.57	42.101
Liver fibrosis	29.31	24.09	3.92	65.8351
Spleen normal	27.41	21.41	4.71	38.2075
Tonsil normal	25.23	19.32	4.63	40.5262
Lymph node normal	26.15	20.35	4.51	43.8889

Small intestine normal	28.23	21.73	5.22	26.8302
Skin-Decubitus	27.18	22.82	3.06	119.908
Synovium	28	21.12	5.59	20.6889
BM-MNC	26.13	19.32	5.51	21.9445
Activated PBMC	25.2	17.95	5.96	16.12
Neutrophils	24.45	19.5	3.65	79.66
Megakaryocytes	22.5	18.95	2.26	208.772
Erythroid	24.2	21.69	1.23	427.7975

5. Tissue Distribution of Human 53763

A human vessel and tissue panel was tested revealing highest expression of human 53763 mRNA in normal brain cortex, normal hypothalamus, prostate tumor, normal prostate, dorsal root ganglion, and normal breast tissue (see Table 3, below).

TABLE 3.

Tissue Type	Mean	± 2 Mean	ΔΔ Ct	Expression
Artery normal	40	22.41	16.07	0
Aorta diseased	40	22.05	16.44	0
Vein normal	40	19.75	18.74	0
Coronary SMC	35.16	21.86	11.79	0
HUVEC	40	20.41	18.08	0
Hemangioma	40	18.52	19.96	0
Heart normal	39.43	19.55	18.37	0
Heart CHF	40	18.98	19.5	0
Kidney	39.44	19.76	18.16	0
Skeletal Muscle	38.97	21.57	15.89	0
Adipose normal	40	20.09	18.4	0
Pancreas	38.91	20.84	16.56	0
primary osteoblasts	40	19.87	18.61	0
Osteoclasts (diff)	40	17.09	21.4	0
Skin normal	39.59	21.22	16.86	0
Spinal cord normal	31.72	20.14	10.07	0.9303
Brain Cortex normal	23.07	21.56	0.01	996.5403
Brain Hypothalamus normal	26.15	20.98	3.65	79.3844
Nerve	39.08	21.23	16.33	0
DRG (Dorsal Root Ganglion)	31.66	21.3	8.86	2.1596
Breast normal	27.25	20.41	5.33	24.9468
Breast tumor	40	20.02	18.46	0
Ovary normal	40	19.66	18.83	0
Ovary Tumor	40	19.7	18.79	0
Prostate Normal	29.68	19.32	8.85	2.1671
Prostate Tumor	28.14	19.95	6.67	9.8204
Salivary glands	40	18.97	19.52	0
Colon normal	39.09	17.8	19.78	0
Colon Tumor	40	18.63	19.86	0
Lung normal	40	17.49	21	0
Lung tumor	39.66	19.81	18.34	0
Lung COPD	40	17.97	20.52	0
Colon IBD	40	17.3	21.18	0

Liver normal	40	19.57	18.91	0
Liver fibrosis	40	21.34	17.15	0
Spleen normal	40	19.27	19.22	0
Tonsil normal	33.5	16.75	15.24	0.0258
Lymph node normal	38.61	18.4	18.7	0
Small intestine normal	36.56	19.96	15.08	0
Skin-Decubitus	39.43	20.41	17.51	0
Synovium	40	19.32	19.16	0
BM-MNC	40	18.21	20.27	0
Activated PBMC	38.88	17.5	19.88	0
Neutrophils	40	18.38	20.11	0
Megakaryocytes	40	18.09	20.39	0
Erythroid	40	21.23	17.25	0

Tissue Distribution of Human 67076

A human vessel panel was tested revealing highest expression of human 67076 mRNA in normal aorta, diseased artery, and static HUVEC (see Table 4, below).

TABLE 4.

Tissue Type	Mean	± 2 Mean	ΔΔ Ct	Expression
Aortic SMC	25.58	21.16	4.42	46.8762
Coronary SMC	29.11	24.36	4.76	36.906
Huvec Static	23.55	20.59	2.96	128.0696
Huvec LSS	23.41	20.06	3.35	98.073
H/Adipose/MET 8	27.7	20.51	7.18	6.8723
H/Artery/Normal/Carotid/CLN 595	26.82	19.34	7.48	5.6014
H/Artery/Normal/Carotid/CLN 598	28.79	20.41	8.37	3.0226
H/Artery/normal/NDR 352	29.41	21.68	7.73	4.7102
H/IM Artery/Normal/AMC 73	32.65	23.77	8.88	2.1152
H/Muscular Artery/Normal/AMC 236	29.2	23.34	5.87	17.1577
H/Muscular Artery/Normal/AMC 254/	29.68	22.56	7.13	7.1393
H/Muscular Artery/Normal/AMC 259	29.63	22.25	7.37	6.0452
H/Muscular Artery/Normal/AMC 261	30.12	22.67	7.45	5.7191
H/Muscular Artery/Normal/AMC 275	30.2	24.2	6	15.6792
H/Aorta/Diseased/PIT 732	30.73	22.36	8.38	3.0121
H/Aorta/Diseased/PIT 710	29.6	23.14	6.46	11.3199
H/Aorta/Diseased/PIT 711	29.35	22.63	6.72	9.4531
H/Aorta/Diseased/PIT 712	28.77	22.02	6.75	9.2585
H/Artery/Diseased/Iliac/NDR 753	26.11	19.41	6.71	9.585
H/Artery/Diseased/Tibial/PIT 679	29.82	20.34	9.47	1.4101
H/Vein/Normal/SaphenousAMC 107	31.66	21.07	10.59	0.6488
H/Vein/Normal/NDR 239	33.13	21.65	11.49	0.3477
H/Vein/Normal/Saphenous/NDR 237	29.71	20.59	9.12	1.7972
H/Vein/Normal/PIT 1010	28.34	22.05	6.3	12.6914
H/Vein/Normal/AMC 191	28.64	22.15	6.49	11.164
H/Vein/Normal/AMC 130	27.41	21.27	6.14	14.1309
H/Vein/Normal/AMC 188	30.56	24.09	6.46	11.3199
H/Vein/Normal/AMC 196	29.89	20.93	8.96	2.008
H/Vein/Normal/AMC 211	32.55	23.52	9.03	1.9196
H/Vein/Normal/AMC 214	30.93	22.99	7.95	4.058
M/Artery/Diseased/CAR 1174	24.56	23.05	1.5	352.3302
M/Artery/Diseased/CAR 1175	24.98	19.89	5.09	29.2585

M/Aorta/Normal/PRI 286	25.52	18.68	6.84	8.7288
M/Artery/Normal/PRI 324	25.13	20.65	4.48	44.8111
M/Aorta/Normal/PRI 264	24.14	24.74	-0.6	1515.7166
M/Artery/Normal/PRI 320	24.93	20.29	4.64	40.1071
M/Vein/Normal/PRI 328	26.67	20.04	6.63	10.0965
HUVEC Vehicle	26.64	21	5.63	20.1232
HUVEC Mev	25.54	20.3	5.25	26.3692
HAEC Vehicle	25.7	20.66	5.04	30.2903
HAEC Mev	27.84	22.41	5.43	23.1957

Tissue Distribution of Human 67102

- 5 A human tissue panel was tested revealing highest expression of human 67102 mRNA in normal kidney tissue and diseased artery (see Table 5, below).

TABLE 5.

Tissue Type Mean \pm 2 Mean $\Delta\Delta$ Ct Expression

Aortic SMC	28.34	21.88	6.47	11.2807
Coronary SMC	29.98	23.11	6.88	8.5196
Huvec Static	27.55	21.41	6.14	14.18
Huvec LSS	27.72	21.12	6.59	10.3444
H/Adipose/MET 8	30.56	20.57	9.99	0.9834
H/Artery/Normal/Carotid/CLN 595	31.42	20.3	11.13	0.4478
H/Artery/Normal/Carotid/CLN 598	32.23	21.69	10.54	0.6717
H/Artery/normal/NDR 352	31.34	22.44	8.9	2.0933
H/IM Artery/Normal/AMC 73	33.46	23.98	9.48	1.4003
H/Muscular Artery/Normal/AMC 236	30.48	23.52	6.96	8.0321
H/Muscular Artery/Normal/AMC 247	33.9	24.07	9.82	1.1025
H/Muscular Artery/Normal/AMC 254/	31.12	23.43	7.68	4.8594
H/Muscular Artery/Normal/AMC 259	30.47	23.07	7.4	5.9208
H/Muscular Artery/Normal/AMC 261	31.32	22.92	8.4	2.9501
H/Muscular Artery/Normal/AMC 275	31.31	24.78	6.53	10.8212
H/Aorta/Diseased/PIT 732	31.73	22.76	8.97	1.9942
H/Aorta/Diseased/PIT 710	30.33	23.36	6.97	7.9767
H/Aorta/Diseased/PIT 711	31.02	23.3	7.72	4.7265
H/Aorta/Diseased/PIT 712	30.57	22.71	7.86	4.3043
H/Artery/Diseased/Iliac/NDR 753	27.22	20.07	7.15	7.041
H/Artery/Diseased/Tibial/PIT 679	32	21.19	10.81	0.557
H/Vein/Normal/SaphenousAMC 107	31.57	22.08	9.49	1.3859
H/Vein/Normal/NDR 239	33.44	22.16	11.28	0.4021
H/Vein/Normal/Saphenous/NDR 237	31.32	21.01	10.31	0.7877
H/Vein/Normal/PIT 1010	29.86	22.36	7.5	5.5243
H/Vein/Normal/AMC 191	30.36	22.53	7.84	4.3796
H/Vein/Normal/AMC 130	30.08	22.32	7.75	4.6293
H/Vein/Normal/AMC 188	32.93	25.01	7.92	4.129
H/Vein/Normal/AMC 196	32.24	21.61	10.64	0.6288
H/Vein/Normal/AMC 211	36.16	23.51	12.65	0
H/Vein/Normal/AMC 214	35.59	24	11.6	0
M/Artery/Diseased/CAR 1175	29.73	21.84	7.89	4.2011
M/Aorta/Normal /543	34.84	29.17	5.67	19.6408
M/Artery/Diseased/CAR 1174	31.11	26.59	4.52	43.5857

M/Pancreas/PRI 2	32.48	26.33	6.15	14.082
M/Kidney/Normal/MPI 88	30.23	26.84	3.38	96.0547
M/Kidney/Normal/MPI 282	29.34	25.94	3.4	95.0612
HUVEC Vehicle	29.25	21.45	7.8	4.4871
HUVEC Mev	28.16	20.87	7.29	6.3899
HAEC Vehicle	28.14	21.16	6.97	7.9491
HAEC Mev	29.61	22.66	6.95	8.088

In addition, a human vessel panel was tested, which revealed high expression of human 67102 mRNA in normal artery, HUVEC, coronary smooth muscle cells, diseased aorta, and normal hypothalamus (see Table, 6, below).

TABLE 6.

Tissue Type	Mean	± 2 Mean	ΔΔ Ct	Expression
Artery normal	27.32	21.75	5.57	21.0505
Aorta diseased	28.27	21.71	6.55	10.6353
Vein normal	30.38	19.83	10.55	0.6693
Coronary SMC	28.61	22.23	6.38	12.0485
HUVEC	26.32	20.32	6	15.5709
Hemangioma	25.91	19.07	6.83	8.7895
Heart normal	27.16	19.98	7.17	6.9441
Heart CHF	27.2	19.06	8.14	3.545
Kidney	25.54	19.59	5.96	16.12
Skeletal Muscle	30.52	21.5	9.03	1.9196
Adipose normal	30.11	19.95	10.15	0.8771
Pancreas	29.57	21.23	8.33	3.1076
primary osteoblasts	28.09	19.85	8.23	3.3191
Osteoclasts (diff)	29.79	17.02	12.77	0.1432
Skin normal	29.31	21.41	7.89	4.2011
Spinal cord normal	28.3	20.36	7.93	4.0863
Brain Cortex normal	28.25	22.04	6.21	13.5084
Brain Hypothalamus normal	28.93	21.49	7.44	5.7589
Nerve	28.34	21.3	7.04	7.5989
DRG (Dorsal Root Ganglion)	29.16	21.11	8.04	3.7994
Breast normal	27.81	20.47	7.34	6.1508
Breast tumor	29.08	20.41	8.68	2.4466
Ovary normal	26.44	19.7	6.74	9.3878
Ovary Tumor	30.93	19.6	11.34	0.3871
Prostate Normal	28.11	19.48	8.63	2.5241
Prostate Tumor	27.68	19.68	8	3.9063
Salivary glands	28.9	19.18	9.71	1.194
Colon Tumor	27.98	18.82	9.16	1.742
Lung normal	26.96	17.4	9.56	1.3202
Lung tumor	27.82	19.64	8.19	3.4361
Lung COPD	26.38	17.66	8.72	2.3633
Colon IBD	28.27	17.29	10.98	0.4934
Liver normal	29.14	19.58	9.56	1.3248
Liver fibrosis	29.89	21.08	8.8	2.2358
Spleen normal	26.95	19.09	7.86	4.3193
Tonsil normal	25.01	16.8	8.21	3.3654
Lymph node normal	26.3	18.22	8.09	3.6828
Small intestine normal	29.03	19.59	9.45	1.4347

Skin-Decubitus	27.66	20.32	7.34	6.1722
Synovium	28.22	19.23	8.98	1.9804
BM-MNC	29.57	18.46	11.12	0.4509
Activated PBMC	28.38	17.25	11.14	0.4447
Neutrophils	27.43	18.4	9.04	1.8997
Megakaryocytes	26.72	17.88	8.84	2.1822
Erythroid	31.52	21.26	10.26	0.8183
Colon normal	30.07	19.25	10.82	0.5551

Tissue Distribution of Human 44181

5. A human vessel panel was tested revealing highest expression of human 44181 mRNA in LSS HUVEC (see Table 7, below).

TABLE 7

Tissue Type	Mean	B 2 Mean	ΔΔ Ct	Expression
Static Huvec	25.37	19.18	6.19	13.697
LSS Huvec	25.7	20.02	5.68	19.4377
Aortic SMC	28.75	20.32	8.43	2.9095
Coronary SMC	28.52	21.2	7.31	6.3019
H/Adipose/MET 9	36.07	18.41	17.66	0
Diseased Heart /PIT 1	29.28	21.15	8.13	3.5697
H/Artery/Normal/Carotid/CLN 595	37.9	18.32	19.59	0
H/Artery/Normal/Carotid/CLN 598	39.97	19.49	20.48	0
H/Artery/normal/NDR 352	40	20.2	19.8	0
H/Artery/Normal/AMC 150	40	22.27	17.73	0
H/Artery/Normal/AMC 73	40	23.84	16.16	0
IMA / AMC 247	39.73	22.79	16.95	0
IMA / AMC 254	33.79	22.23	11.56	0.3324
IMA / AMC 259	33.68	21.12	12.56	0.1656
IMA / AMC 261	34.73	21.23	13.5	0.0863
IMA / AMC 275	40	24.52	15.48	0
IMA / AMC 279	30.89	22.41	8.48	0
H/Artery/Diseased/Iliac/NDR 753	36.59	18.43	18.16	0
H/Artery/Diseased/Tibial/PIT 679	40	19.84	20.16	0
Aorta/Diseased/PIT 732	34.74	21.32	13.41	0.0916
Aorta/Diseased/ PIT 710	33.04	22.48	10.56	0.6624
Aorta/Diseased/PIT 711	31.89	22.09	9.8	1.1218
Aorta/Diseased/ PIT 712	32.92	22.09	10.84	0.5474
H/Vein/Normal/Saphenous/N DR 721	32.66	16.82	15.83	0.0172
H/Vein/Normal/SaphenousA MC 107	40	20	20	0
H/Vein/Normal/NDR 239	40	20.61	19.39	0
H/Vein/Normal/Saphenous/N DR 237	40	19.1	20.9	0
H/Vein/Normal/NDR 235	40	21.34	18.66	0

H/Vein/Normal/MPI 1101	33.56	19.59	13.98	0.0621
HMVEC/Vehicle/24 hr	30.04	17.84	12.2	0.2125
HMVEC/Mev/24hr/1X	29.77	18	11.76	0.2883
HMVEC/MEV/24HR/2.5X	30.32	18.67	11.65	0.3112
HMVEC/MEV/48HR/1X	31.58	18.8	12.79	0.1417
HMVEC/MEV/48HR/2.5X	31.77	18.37	13.4	0.0922
HUVEC/Vehicle/24 hr	30.5	18.15	12.36	0.1909
HUVEC/Mev/24hr/1X	30.28	17.52	12.76	0.1442
HUVEC/MEV/24HR/2.5X	29.35	19.18	10.18	0.865
HUVEC/MEV/48HR/1X	35.68	21.54	14.14	0
HUVEC/MEV/48HR/2.5X	34.7	23	11.7	0.3016

Tissue Distribution of Human 67084

- 5 A human vessel panel was tested revealing highest expression of human 67084 mRNA in HUVEC, LSS HUVEC, and coronary smooth muscle cells (see Table 8, below).

TABLE 8.

Tissue Type	Mean	± 2 Mean	ΔΔ Ct	Expression
Aortic SMC	25.92	19.23	6.7	9.6517
Coronary SMC	26.59	20.36	6.23	13.3224
Huvec Static	23.39	18.5	4.88	33.843
Huvec LSS	24.31	18.32	5.99	15.7883
H/Adipose/MET 9	26.4	18.46	7.94	4.0721
H/Artery/Normal/Carotid/CLN 595	26.83	18.84	8	3.9198
H/Artery/Normal/Carotid/CLN 598	28.49	20.16	8.34	3.0968
H/Artery/normal/NDR 352	27.12	20.32	6.8	8.9432
H/IM Artery/Normal/AMC 73	31.48	23.36	8.12	3.607
H/Muscular Artery/Normal/AMC 236	30.93	23.56	7.38	6.0243
H/Muscular Artery/Normal/AMC 247	33.77	24.84	8.92	2.0645
H/Muscular Artery/Normal/AMC 254/	30.69	23.68	7	7.7855
H/Muscular Artery/Normal/AMC 259	29.9	22.12	7.78	4.5497
H/Muscular Artery/Normal/AMC 261	29.93	21.13	8.8	2.2436
H/Muscular Artery/Normal/AMC 275	30.29	22.97	7.33	6.2367
H/Aorta/Diseased/PIT 732	29.02	21.35	7.67	4.8932
H/Aorta/Diseased/PIT 710	31.36	22.8	8.56	2.6496
H/Aorta/Diseased/PIT 711	31.31	22.6	8.71	2.3963
H/Aorta/Diseased/PIT 712	31.4	22.48	8.92	2.0645
H/Artery/Diseased/Iliac/NDR 753	25.37	17.73	7.64	4.996
H/Artery/Diseased/Tibial/PIT 679	28.55	19.45	9.11	1.816
H/Vein/Normal/SaphenousAMC 107	29.48	21.11	8.38	3.0121
H/Vein/Normal/Saphenous/NDR 237	28.67	19.86	8.8	2.2358
H/Vein/Normal/PIT 1010	28.31	20.55	7.76	4.5973
H/Vein/Normal/AMC 191	29.25	20.77	8.47	2.8104
H/Vein/Normal/AMC 130	28.32	20.45	7.88	4.2598
H/Vein/Normal/AMC 188	31.68	24.61	7.06	7.4943
H/Vein/Normal/NDR 239	35.65	29.23	6.42	0
HUVEC Vehicle	26.86	20.14	6.71	9.5188
HUVEC Mev	25.83	18.52	7.3	6.3238
HAEC Vehicle	26.57	19.64	6.94	8.1443
HAEC Mev	27.81	21.13	6.67	9.7864

EXAMPLE 3: EXPRESSION OF RECOMBINANT 67076, 67102, 44181, 67084FL, or 67084alt POLYPEPTIDE IN BACTERIAL CELLS

5 In this example, human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the
10 GST-PLTR fusion polypeptide in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

15

EXAMPLE 4: EXPRESSION OF RECOMBINANT 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt POLYPEPTIDES IN COS CELLS

20 To express the human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire
25 PLTR polypeptide and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant polypeptide under the control of the CMV promoter.

To construct the plasmid, the human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt DNA sequence is amplified by PCR using two primers. The
30 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon,
35 the HA tag or FLAG tag and the last 20 nucleotides of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt coding sequence. The PCR amplified fragment and the pcDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA).

Preferably the two restriction sites chosen are different so that the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5 α , SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the IC54420 polypeptide is detected by radiolabelling (^{35}S -methionine or ^{35}S -cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with ^{35}S -methionine (or ^{35}S -cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the human 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt polypeptide is detected by radiolabelling and immunoprecipitation using a 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt -specific monoclonal antibody.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:
 - 5 (a) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:7, SEQ ID NO:10, or SEQ ID NO:13, SEQ ID NO:16, SEQ ID NO:19, SEQ ID NO:22, or SEQ ID NO:25; and
 - (b) a nucleic acid molecule comprising the nucleotide sequence set forth in
10 SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:9, SEQ ID NO:12, or SEQ ID NO:15, SEQ ID NO:18, SEQ ID NO:21, SEQ ID NO:24, or SEQ ID NO:27.
2. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID
15 NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26.
3. An isolated nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____, _____, _____, or _____.
- 20 4. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26.
- 25 5. An isolated nucleic acid molecule selected from the group consisting of:
 - a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID
30 NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or a complement thereof;
 - b) a nucleic acid molecule comprising a fragment of at least 30
35 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID

NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27, or a complement thereof;

5 c) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 60% identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26; and

10 d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, wherein the fragment comprises at least 10 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26.

15 6. An isolated nucleic acid molecule which hybridizes to a complement of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 under stringent conditions.

20 7. An isolated nucleic acid molecule comprising a nucleotide sequence which is complementary to the nucleotide sequence of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

8. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5, and a nucleotide sequence encoding a heterologous polypeptide.

25 9. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

10. The vector of claim 9, which is an expression vector.

30 11. A host cell transfected with the expression vector of claim 10.

12. A method of producing a polypeptide comprising culturing the host cell of claim 11 in an appropriate culture medium to, thereby, produce the polypeptide.

35 13. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID

NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26;

5 b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27 under stringent conditions;

10 c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:27; and

20 d) a polypeptide comprising an amino acid sequence which is at least 60% identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26.

14. The isolated polypeptide of claim 13 comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:11, SEQ ID NO:14, SEQ ID NO:17, SEQ ID NO:20, SEQ ID NO:23, or SEQ ID NO:26.

15. The polypeptide of claim 13, further comprising heterologous amino acid sequences.

30

16. An antibody which selectively binds to a polypeptide of claim 13.

17. A method for detecting the presence of a polypeptide of claim 13 in a sample comprising:

35 a) contacting the sample with a compound which selectively binds to the polypeptide; and

 b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 13 in the sample.

18. The method of claim 17, wherein the compound which binds to the polypeptide is an antibody.

5 19. A kit comprising a compound which selectively binds to a polypeptide of claim 13 and instructions for use.

20. A method for detecting the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in a sample comprising:

- 10 a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to a complement of the nucleic acid molecule; and
b) determining whether the nucleic acid probe or primer binds to the complement of the nucleic acid molecule in the sample to thereby detect the presence of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in the sample.

15 21. The method of claim 20, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

22. A kit comprising a compound which selectively hybridizes to a complement of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 and instructions for use.

23. A method for identifying a compound which binds to a polypeptide of claim 13 comprising:

- 25 a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and
b) determining whether the polypeptide binds to the test compound.

24. The method of claim 23, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

- 30 a) detection of binding by direct detection of test compound/polypeptide binding;
b) detection of binding using a competition binding assay; and
c) detection of binding using an assay for 8099, 46455, 54414, 53763, 67076, 67102, 44181, 67084FL, or 67084alt activity.

35 25. A method for modulating the activity of a polypeptide of claim 13 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound

which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

26. A method for identifying a compound which modulates the activity of a polypeptide of claim 13 comprising:

- a) contacting a polypeptide of claim 13 with a test compound; and
- b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

10

CCACGCGTCCGGCCTTCCGAAATAGAAA^UAAAGTTGGTCA^UAAATCACATTAGCTTTGCCCGAAGTTTTC^UCCCACACT
 CTTCTTTAGCATGCTATTATGGGGA^UAGTGACCACTCCTGGGAGCGGGGGTGGT^UCGGGCGGTTTGGTGGCGGGGAAGC

M V P V E N T E G P S L L N 14
 GGCTGTA^UACTTCTACGTGACC ATG. GTA CCT GTT GAA. AAC ACC GAG GGC CCC AGT CTG CTG AAC 42

Q K G T A V E T E G S G S R H P P W A R 34
 CAG AAG GGG ACA GCC GTG GAG ACG GAG GGC AGC GGC AGC CGG CAT CCT CCC TGG GCG AGA 102

G C G M F T F L S S V T A A V S G L L V 54
 GGC TGC GGC ATG TTT ACC TTC CTG TCA TCT GTC ACT GCT GCT GTC AGT GGC CTC CTG GTG 162

G Y E L G I I S G A L L Q I K T L L A L 74
 GGT TAT GAA CTT GGG ATC ATC TCT GGG GCT CTT CTT CAG ATC AAA ACC TTA TTA GCC CTG 222

S C H E Q E M V V S S L V I G A L L A S 94
 AGC TGC CAT GAG CAG GAA ATG GTT GTG AGC TCC CTC GTC ATT GGA GCC CTC CTT GCC TCA 282

L T G G V L I D R Y G R R T A I I L S S 114
 CTC ACC GGA GGG GTC CTG ATA GAC AGA TAT GGA. AGA AGG ACA GCA ATC ATC TTG TCA TCC 342

C L L G L G S L V L I L S L S Y T V L I 134
 TGC CTG CTT GGA CTC GGA AGC TTA GTC TTG ATC CTC AGT TTA TCC TAC ACG GTT CTT ATA 402

V G R I A I G V S I S L S S I A T C V Y 154
 GTG GGA CGC ATT GCC ATA GGG GTC TCC ATC TCC CTC TCT TCC ATT GCC ACT TGT GTT TAC 462

I A E I A P Q H R R G L L V S L N E L M 174
 ATC GCA GAG ATT GCT CCT CAA CAC AGA AGA GGC CTT CTT GTG TCA CTG AAT GAG CTG ATG 522

I V I G I L S A Y I S N Y A F A N V F H 194
 ATT GTC ATC GGC ATT CTT TCT GCC TAT ATT TCA AAT TAC GCA TTT GCC AAT GTT TTC CAT 582

G W K Y M F G L V I P L G V L Q A I A M 214
 GGC TGG AAG TAC ATG TTT GGT CTT GTG ATT CCC TTG GGA GTT TTG CAA GCA ATT GCA ATG 642

Y F L P P S P R F L V M K G Q E G A A S 234
 TAT TTT CTT CCT CCA AGC CCT CGG TTT CTG GTG ATG AAA GGA CAA GAG GGA GCT GCT AGC 702

K V L G R L R A L S D T T E E L T V I K 254
 AAG GTT CTT GGA AGG TTA AGA GCA CTC TCA GAT ACA ACT GAG GAA CTC ACT GTG ATC AAA 762

S S L K D E Y Q Y S F W D L F R S K D N 274
 TCC TCC CTG AAA GAT GAA TAT CAG TAC AGT TTT TGG GAT CTG TTT CGT TCA AAA GAC AAC 822

M R T R I M I G L T L V F F V Q I T G Q 294
 ATG CGG ACC CGA ATA ATG ATA GGA CTA ACA CTA GTA TTT TTT GTA CAA ATC ACT GGC CAA 882

P N I L F Y A S T V L K S V G F Q S N E 314
 CCA AAC ATA TTG TTC TAT GCA TCA ACT GTT TTG AAG TCA GTT GGA TTT CAA AGC AAT GAG 942

A A S L A S T G V G V V K V I S T I P A 334
 GCA GCT AGC CTC GCC TCC ACT GGG GTT GGA GTC GTC AAG GTC ATT AGC ACC ATC CCT GCC 1002

Figure 1A

T	L	L	V	D		V	G	S	K	T	F	L	C	I	G	S	S	V	M	354
ACT	CTT	CTT	GTA	GAC	CAT	G	GGC	AGC	AAA	ACA	TTC	CTC	TC	TT	GG	TCC	TCT	GTG	ATG	1062
A	A	S	L	V	T	M	G	I	V	N	L	N	I	H	M	N	F	T	H	374
GCA	GCT	TCG	TTG	GTG	ACC	ATG	GGC	ATC	GTA	AAT	CTC	AAC	ATC	CAC	ATG	AAC	TTC	ACC	CAT	1122
I	C	R	S	H	N	S	I	N	Q	S	L	D	E	S	V	I	Y	G	P	394
ATC	TGC	AGA	AGC	CAC	AAT	TCT	ATC	AAC	CAG	TCC	TTG	GAT	GAG	TCT	GTG	ATT	TAT	GGA	CCA	1182
G	N	L	S	T	N	N	N	T	L	R	D	H	F	K	G	I	S	S	H	414
GGA	AAC	CTG	TCA	ACC	AAC	AAC	AAT	ACT	CTC	AGA	GAC	CAC	TTC	AAA	GGG	ATT	TCT	TCC	CAT	1242
S	R	S	S	L	M	P	L	R	N	D	V	D	K	R	G	E	T	T	S	434
AGC	AGA	AGC	TCA	CTC	ATG	CCC	CTG	AGA	AAT	GAT	GTG	GAT	AAG	AGA	GGG	GAG	ACG	ACC	TCA	1302
A	S	L	L	N	A	G	L	S	H	T	E	Y	Q	I	V	T	D	P	G	454
GCA	TCC	TTG	CTA	AAT	GCT	GGA	TTA	AGC	CAC	ACT	GAA	TAC	CAG	ATA	GTC	ACA	GAC	CCT	GGG	1362
D	V	P	A	F	L	K	W	L	S	L	A	S	L	L	V	Y	V	A	A	474
GAC	GTC	CCA	GCT	TTT	TTG	AAA	TGG	CTG	TCC	TTA	GCC	AGC	TTG	CTT	GTT	TAT	GTT	GCT	GCT	1422
F	S	I	G	L	G	P	M	P	W	L	V	L	S	E	I	F	P	G	G	494
TTT	TCA	ATT	GGT	CTA	GGA	CCA	ATG	CCC	TGG	CTG	GTG	CTC	AGC	GAG	ATC	TTT	CCT	GGT	GGG	1482
I	R	G	R	A	M	A	L	T	S	S	M	N	W	G	I	N	L	L	I	514
ATC	AGA	GGA	CGA	GCC	ATG	GCT	TTA	ACT	TCT	AGC	ATG	AAC	TGG	GGC	ATC	AAT	CTC	CTC	ATC	1542
S	L	T	F	L	T	V	T	D	L	I	G	L	P	W	V	C	F	I	Y	534
TCG	CTG	ACA	TTT	TTG	ACT	GTA	ACT	GAT	CTT	ATT	GGC	CTG	CCA	TGG	GTG	TGC	TTT	ATA	TAT	1602
T	I	M	S	L	A	S	L	L	F	V	V	M	F	I	P	E	T	K	G	554
ACA	ATC	ATG	AGT	CTA	GCA	TCC	CTG	CTT	TTT	GTT	GTT	ATG	TTT	ATA	CCT	GAG	ACA	AAG	GGA	1662
C	S	L	E	Q	I	S	M	E	L	A	K	V	N	Y	V	K	N	N	I	574
TGC	TCT	TTG	GAA	CAA	ATA	TCA	ATG	GAG	CTA	GCA	AAA	GTG	AAC	TAT	GTG	AAA	AAC	AAC	ATT	1722
C	F	M	S	H	H	Q	E	E	L	V	P	K	Q	P	Q	K	R	K	P	594
TGT	TTT	ATG	AGT	CAT	CAC	CAA	GAA	GAA	TTA	GTG	CCA	AAA	CAG	CCT	CAA	AAA	AGA	AAA	CCC	1782
Q	E	Q	L	L	E	C	N	K	L	C	G	R	G	Q	S	R	Q	L	S	614
CAG	GAG	CAG	CTC	TTG	GAG	TGT	AAC	AAG	CTG	TGT	GGT	AGG	GGC	CAA	TCC	AGG	CAG	CTT	TCT	1842
P	E	T	*																	618
CCA	GAG	ACC	TAA																	1854

TGGCCTCAACACCTTCTGAACGTGGATAGTGCCAGAACACTTAGGAGGGTGTCTTTGGACCAATGCATAGTTGCGACTC
CTGTGCTCTCTTTTCACTGTGCATGGAACCTGGTTTTGAAGAGACACTCTGAAATGATAAAGACAGCCTTTAATCCCCCTC
CTCCCCAGAAGGAACCTCAAAAGGTAGATGAGGTACAAGGTCCTAAGTGATCTCTTTTCTGAGCAGGATATCAGGTTA
AAAAAAAAAAGTTACTGGCTGGTTTAATACTTTCTACCTTCTTCACAGAGCAGCCTTTGAATAGACTATGTCCTAGTGA
AGACATCAACCTCCGCCTTAAGCTATGTATGTATGGAGGCCAGTCCGAGCTTTATTATGCAGACACACAAGTGGTCTGG
ACATGAGGGTACAGTTTCTGCCTACCAAGACACTACTTGCACTGGATCTTACGCAAAAAAGAACCAGAACACACAGTGT
GGACAACAGTCCCATATATTCTATCTAGATTAGGAGAGGGTCTGCTAGGATTTTATGTTAATTCTAGTTACATTCA
ACAAGTATAAAGATTATAGAGCTTATTTTATGAACTATAAACTATAATTTAATGCAAAATATCCTTTTATGAATTTTCAT
GTTAATATTGTGAAATATTAAATAATTCCGCAATAAAAAAAAAAAAAAGGGCGGCCGC

Figure 1B

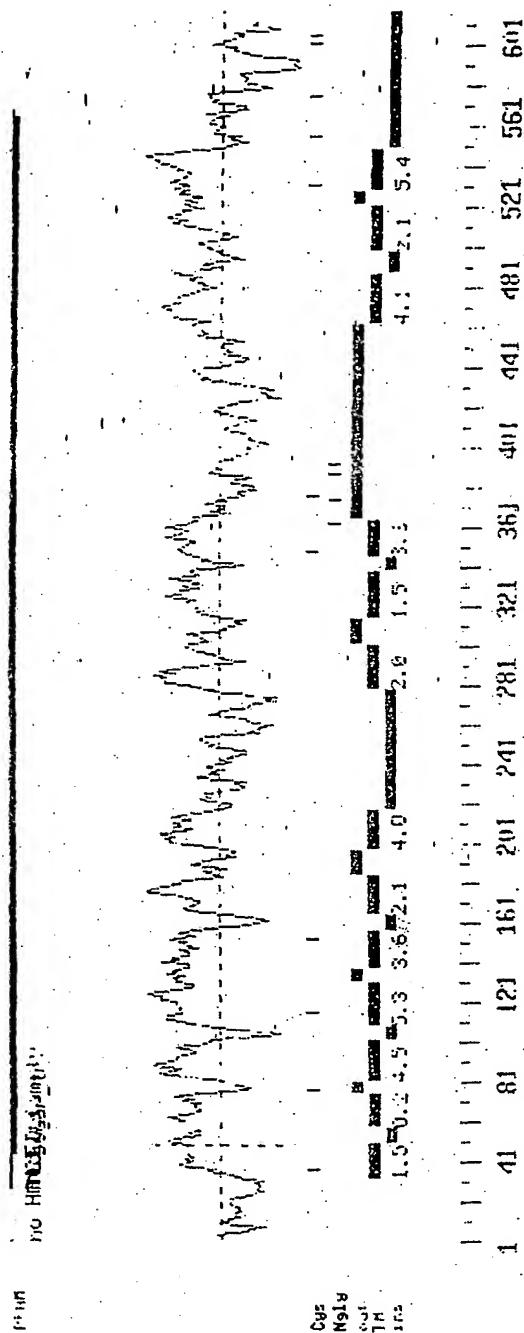


FIGURE 2

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.8089.seq

Query: 8099

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
sugar_tr 1	Sugar (and other) transporter	318.2	9.6e-92	
FecCD_family 1	FecCD transport family	-218.2	6.9	
MCT 1	Monocarboxylate transporter	-235.8	2.7	

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
FecCD_family	1/1	26	227	1	311 []	-218.2	6.9
sugar_tr	1/1	43	564	1	488 []	318.2	9.6e-92
MCT	1/1	29	567	1	611 []	-235.8	2.7

Alignments of top-scoring domains:

FecCD_family: domain 1 of 1, from 26 to 227: score -218.2, E = 6.9

```

      *-->GalsispadvlqalfgggtegeievdeliwdltlrRLPRVLIAlLV
      G+++ ++a  ++  + ++  ++  + +LLV
8099  26  GSRHPPWARGCGMFT---FLSS-----VTAA-----VSGLLV 54

      GAaLAVaGailQglrNPLAsPgilGinsGAslgvvlaivlfpgglsisa
      G      + lGi sGA l +  ++l + + ++
8099  55  G-----YELGIISGALLQIKTLLALSCH-EQEMV 82

      lyllpsfAfaGaliallVyllawkgrnglspvrLiLaGialsalFsAlt
      +l+++A++ +l++++l+  +++  + i+ls+++ +l
8099  83  VSSLVIGALLASLTGGVLIDRYGRR-----TAILSSCHLGLG 120

      tlllllsddlqdgqalfWltGSlsgrnWedvklalpilliglplalllar
      +l.l+ls + +++  + gr  v + l ++ + +a +
8099  121 SLVLILSLSYTVL-----IVGRIAIGVSISSSIATCVYIAEI--- 158

      qLnVLSLGddtAkglGvnvervR.llllllvvlLtGaaVAVAGpIgFVGL
      +++ R+ll++l  +++  +G+
8099  159 -----APQHRRgLLVSLNELMIV-----IGI 179

      ivPHiaRrLvGt.dhrwLLPaSAlGAILLlADllARTlfaPiElPvGi
      + +i  h w  +++++ + +P G+
8099  180 LSAYISNYAFANvFHGW-----KYMFGVLVPLGV 208

      vTaliGaPyFl.....YLLrr<--*
      + A+ a+yFl++++++L+++
8099  209 LQAI--AMYFLppsprFLVMK 227

```

sugar_tr: domain 1 of 1, from 43 to 564: score 318.2, E = 9.6e-92

```

      *-->valvaalgGgflfGyDtgvggflalidflfrfglltssgalaelvg
      +aaa+ G +l Gy +g+i+g+l +i+  l s+ ++
8099  43  SSVTAAVSG-LLVGYELGIISGALLQIK---TLLALSCHQE---- 80

```

FIGURE 3A

BNSDOCID: <WO__02055701A2_1_>

```

8099 213 ----AMYFLP-----PSPRFLVMKGQEGAASKVLGRLRLSPDTEELTV 252
      lpkasiikledakaersvdsLlsskSvgerdksqkseksqasgrpass
8099 253 I----- 253
      atavqlvllrsrlekadlplkrvrsvrrrvlSkVSaeSgtdgersSgylN
      +s+l
8099 254 -----KSSL----- 257
      rkdvFYtGsisNvaeekedpdkYrssslhgtrttvgnaesqstlrlddsr
8099 -----
      esgdgdsssedlsektrgdggkessskeiretikllDfsvlknrtFlI
      k ++ +++++ ++ + + + r ++
8099 258 -----KDEYQYSFWDLFR---SKDNMRTRIMIG 282
      yaisnlfaslGffvPlvflvsYaikslgldekeAsfllsi.iGvsnivGR
      +++ +++ G l++ + ks g ++eA+ L s+++Gv+ +++
8099 283 LTLVFFVQITGQPNILFYASTVL-KSVGFQSNEAASLASTgVGvVKVIST 331
      pifGlVADkkgvrpTarihivifnlsllalGlttlacPlatsfwgLvvyg
      + ++l+ D+ g ++ + +s+++ l t+ + + ++ +c
8099 332 IPATLLVDHVGSKT-----FLCIGSSVMAASLVTMGIVNLNIHMNFTHIC 376
      ilFGfs.....
      +++++ +++ +++++ +++++ +++++ +++++ + +++
8099 377 RSHNSInqsldeSVIYgpgnlstnnntlrdfhkgisshsrsslmplrndv 426
      .....
      +++++++ + + +++++ + +++++ + + + + +
8099 427 dkrgettsasllnaglshteyqivtdpgdvpafkwlsasllvyvaafs 476
      .....iGsygaLtfvvLvdlvg
      + ++ + ++ +++ ++ + + + ++Ltf +++dL+g
8099 477 iglgmpwlvlseifpggirgramaltssmnwgINLLISLTFLTVDLIG 526
      WlekfsnAfGllllfeGvavLvGPPiaGllvDakttgdYtvaFyfsGill
      + + +++++ +a+L ++ + + t+ + ++ +
8099 527 -----LPWVCFIYTIMSLASLLFVVMF---IPE--TKGCSLEQIS----M 562
      llsgl<--+
      l+ +
8099 563 ELAKV 567

```

FIGURE 3C

[illegible]

FIGURE 5

-----TRANSPRTERMVTINTESALT-----PRSLRDTTRMNMVSVAAVAGLLFGLDIG
-----TRANSPRTERMPDAKKQG-----RSNKAMTFFVCFLAALAGLLFGLDIG
MVPVENTEGPSSLNQGTA VETEGSGSRHPPWARGCGMFTLSSVTA VSGLLVGYELGJ
* . . . * . . . * . . . * . . . * . . . * . . . * . . . * . . . *

IAGALPFI TDHFVLTSRLQEWVVSMMGLGAAIGALFNGLSFR LGRKYSLMAGAILFVLC
IAGALPFI ADEFOITSH TQEWVVSMMFGAAVGAVSGWLSFKLGRKKSLMIGAILFVAC
ISGALLQIKTLLALSCHEQEMVVS SLVIGALLASLTGGVLIDRYGRRTAI ILSSCLLGLC
* : * * * : : : : * * * * * : : * * : : : * * : * * : : : * * : * * : : : *

SIGSAFATSVEMLIAARVVLGIAVGIASYTAPLYLSEMASENVRGKMISMYQLMVTLEGIV
SLFSAAAPNVEVLILSRVLLGLAVGVASYTAPLYLSEIAPEKIRGSMISMYQLMITIGIL
SLVLILSLSYTVLIVGRIAIGVSISLSSIATCVYIAEIAPOHRRGLLVSLNELMIVIGIL
*: : : ** *: *: : : : * : : *: : * : : * : : * : : *

LAFSLDTAFS--YSGNWRAMLGVLALP-AVLLIILVVFLPNSPRWLAEKGRHIEAEFVLR
GAYLSDTAFS--YTGAWRWMLGVIIIP-AILLIGVFFLPDSPRWFAAKRRFVDAERVLL
SAYISNYAFANVFHG-WKYMFG-LVPLGLVQAIAMYFLPPSPRFLVMKGQEGAASKVLG
::: **: : * *: *: * : : * : * * * * *: : * : * * *

MLRDTSEKAREELNEIRESLKLK-Q-GGWALFKINRNVRRVFLGMLLQAMQQFTGMNII
RLRDTSAEAKRELDEIRESLOVK-Q-SGWALFKENSFRRVFLGVLLQVMQQFTGMNVI
RLRALSDTT-EELTVIKSSLKDEYQYSFWDLFRSKDNMRTRIMIGLTLVFFVQITGQPN
* * * : * * * : * : * * : * * : * * : * * : *

MYYPRIFKMAGFTTTEQQMIATLVVGLTFMFATFIAVFTVDKAGRKPALKIGFSVMA LG
MYYPKIFELAGYTNTEQQMWGTIVVGLTNVLATFIAIGLVDRWGRKPTLTGLFLVMAAG
LFYASTVLKSVGFQSN EASLASTGVGVVKVISTIPATLLVDHVGSKTFLCIGSSVMAAS
:::*. ::. *. :. :. :*. :. :*: * :*: * * * :* :* :*

TLVLGYCLMQFDN-----GTASSG--
MGVLG-TMMHI-----GIHSPS--
LVTMGIVNLNIHMNFTHICRSHNSINQSLDESVIYGPNLSTNNNTLRDHPFKGISSHSRS
.:* :.: * *

-----LSWLSVGMTMMCIAGYAM
-----AQYFAIAMLLMFIVGFAM
SLMPLRNDVDKRGETTSASLLNAGLSHTEYQIVTDPGDVPAFLKWSLASLLVYVAAFSI
 : : : : : : : : : : :

SAAPVVWILCSEIQP--LKCRDFGITCSTTTNWVSNMIIGATFLTLSDSIGAGTFWLYT
SAGPLIWVLCSEIQP--LKGRDFGITCSTATNWIANMIVGATFLTMLNTLGNANTFWVYA
GLGMPWLVLSEIFPGGIRGRAMALTSS--MNWGINLLISLTLFTVDTLIGLPWVCFIYT
. . *: *:: *** * :: * :.:*. * ** *:.:. ****: : : * :.:: :

ALNIAFVGITFWLIPETKNVTLEHIERKLMAGEKLRLN-----IGV-----
ALNVLFILLTLWLVPETKHVSLEHIERNLMKGKRKLR-----IGAHD-----
IMSLASLLFVVMFIPETKGCSELIQISMELAKVNYVKNNICFMSSHQEELVPKQPQKRKPQ
.: : . :

EQLECNKLCGRGOSROLSPET

9/79

jh8099FL MVPVENTEGPSLLNQKGTAVETEGSG-- --SRHPPWARG-CGMFTFLSSVTAASVGLLVGY
2168Patent ---MTSDHEHMTAVCASHVQTHGSOLQIQKLSPCFRPPTPAFRISSIILLGAG-LAGP
.: : .: *:*.*. :. * * *: **: :* **

jh8099FL ELGIISGALLQIKTLLALSCHQEEMVVSSLVIGALLASLTGGVLIDRYGRRTAILSSCL
2168Patent STGDRWFGVSUVGTGLFPLPQLLLPRLLFTHAILERLHLWLALPPVLVLGHALLH-CK
. * .: : * * *. : : *.* * : : : *

jh8099FL LGLGSLVLILSLSYTVLIVGRIAIGVISLSSIATCVYIAEIAPQH--RRGLLVSLNELM
2168Patent VGGSTARAGDQLVQRVLLL-IVFLHRWVQVPZGTEVDILMGMSRTGGRRGPPELRP---G
:* : . * **: : : : : . * * * :... *** :

jh8099FL IVIGILSAYISNYAFANVFHGWKYMFGGLVIPLGVLOAIAMYFLPPSPRFLVMKGQEGAAS
2168Patent FRISILSAYISNYAFANVFHGWKYMFGGLVIPLGVLOAIAMYFLPPSPRFLVMKGQEGAAS
: * *****

jh8099FL KVLGRLRALSDTTEELTVIKSSLKDEYQYSFWDLFRSKDNMRTRIMIGTLVFFVQITGO
2168Patent KVLGRLRALSDTTEELTVIKSSLKDEYQYSFWDLFRSKDNMRTRIMIGTLVFFVQITGO

jh8099FL PNILFYASTVLKS SVGFQSNEAASLASTGVG VVKVISTI PATLLVDHVGSKTFLCIGSSVM
2168Patent PNILFYASTVLKS SVGFQSNEAASLASTGVG VVKVISTI PATLLVDHVGSKTFLCIG---

jh8099FL AASLVTMGI VNLNIHMNFTHICRS HNSINQSLDES VIYGPNLSTNNNTLRDHFKGIS SH
2168Patent -----

jh8099FL SRSSLMPLRNDVDKRGETTSASLLNAGLSHTHEYQIVTDPGDVPAFLKWLSLASLLVYVAA
2168Patent -----LLNAGLSHTHEYQIVTDPGDVPAFLKWLSLASLLVYVAA

jh8099FL FSIGLGMPWLVLSEIFPGGIRGRAMALTSSMNWGINLLISLTFLTVDLIGLPWVCFIY
2168Patent FSIGLGMPWLVLSEIFPGGIRGRAMALTSSMNWGINLLISLTFLTVDLIGLPWVCFIY

jh8099FL TIMSLASLLFVVMFIPETKGC SLEQISMELAKVNYVKNNICFMSSHQEELVPKPQKRKP
2168Patent TIMSLASLLFVVMFIPETKGC SLEQISMELAKVNYVKNNICFMSSHQEELVPKPQKRKP

jh8099FL QEOLLEC NKLCGRGQSRQLSPET
2168Patent QEOLLEC NKLCGRGQSRQLSPET

FIGURE 7

Input file Fbh46455FL.seq Output File Fbh46455FL.cra
Sequence length 2230

GTCTGACCCACGCGTCCGGCAACATGGCGGCTGCCGTGGTGCAGCGCCCGGGCTGAGCGACAGCAAGTGCAGCGGGCTCC
TACCCCGGGTGAGGGGTGGCCCTCCGCGTGCGATCGTGCCCTCTTCAGCCCGCTCCTGTCCCCGACATCACGTGTATTCC
GCACGTCCCCCTCCGCGCTGTGTGTCTACTGAGACGGGGAGGCGTGACAGGGCCCCGGGTCCCTTCTCAGTGGTCTCTGT
GCTTCAGGGCAAGCTCCCCGTCTCCGGGCGCACTTCCCTCGCCTGTGTTCCGGTCCATCCTCTTCTCCAGCCTCCTCC
CCTCGCAGGTGGGATCGTGGTGGGACCGGAGCGCGGGCGGGCGGGCCCGGGGACCAATG GCC GGG TCC GAC 5
15
T A P F L S Q A D D P D D G P V P G T P 25
ACC GCG CCC TTC CTC AGC CAG GCG GAT GAC CCG GAC GAC GGG CCA GTG CCT GGC ACC CCG 75
G L P G S T G N P K S E E P E V P D Q E 45
GGG TTG CCA GGG TCC ACG GGG AAC CCG AAG TCC GAG GAG CCC GAG GTC CCG GAC CAG GAG 135
G L Q R I T G L S P G R S A L I V A V L 65
GGG CTG CAG CGC ATC ACC GGC CTG TCT CCC GGC CGT TCG GCT CTC ATA GTG GCG GTG CTG 195
C Y I N L L N Y M D R F T V A G V L P D 85
TGC TAC ATC AAT CTC CTG AAC TAC ATG GAC CGC TTC ACC GTG GCT GGC GTC CTT CCC GAC 255
I E Q F F N I G D S S S G L I Q T V F I 105
ATC GAG CAG TTC TTC AAC ATC GGG GAC AGT AGC TCT GGG CTC ATC CAG ACC GTG TTC ATC 315
S S Y M V L A P V F G Y L G D R Y N R K 125
TCC AGT TAC ATG GTG TTG GCA CCT GTG TTT GGC TAC CTG GGT GAC AGG TAC AAT CGG AAG 375
Y L M C G G I A F W S L V T L G S S F I 145
TAT CTC ATG TGC GGG GGC ATT GCC TTC TGG TCC CTG GTG ACA CTG GGG TCA TCC TTC ATC 435
P G E H F W L L L L T R G L V G V G E A 165
CCC GGA GAG CAT TTC TGG CTG CTC CTC CTG ACC CGG GGC CTG GTG GGG GTC GGG GAG GCC 495
S Y S T I A P T L I A D L F V A D Q R S 185
AGT TAT TCC ACC ATC GCG CCC ACT CTC ATT GCC GAC CTC TTT GTG GCC GAC CAG CGG AGC 555
R M L S I F Y F A I P V G S G L G Y I A 205
CGG ATG CTC AGC ATC TTC TAC TTT GCC ATT CCG GTG GGC AGT GGT CTG GGC TAC ATT GCA 615
G S K V K D M A G D W H W A L R V T P G 225
GGC TCC AAA GTG AAG GAT ATG GCT GGA GAC TGG CAC TGG GCT CTG AGG GTG ACA CCG GGT 675
L G V V A V L L L F L V V R E P P R G A 245
CTA GGA GTG GTG GCC GTT CTG CTG CTG TTC CTG GTA GTG CGG GAG CCG CCA AGG GGA GCC 735
V E R H S D L P P L N P T S W W A D L R 265
GTG GAG CGC CAC TCA GAT TTG CCA CCC CTG AAC CCC ACC TCG TGG TGG GCA GAT CTG AGG 795
A L A R N P S F V L S S L G F T A V A F 285
GCT CTG GCA AGA AAT CCT AGT TTC GTC CTG TCT TCC CTG GGC TTC ACT GCT GTG GCC TTT 855
V T G S L A L W A P A F L L R S R V V L 305
GTC ACG GGC TCC CTG GCT CTG TGG GCT CCG GCA TTC CTG CTG CGT TCC CGC GTG GTC CTT 915

FIGURE 8A

G E T P P C L P G D S C S S S D S L I F	325
GGG GAG ACC CCA CCC TGC CTT CCC GGA GAC TCC TGC TCT TCC TCT GAC AGT CTC ATC TTT	975
G L I T C L T G IV L G V G L G V E I S R	345
GGA CTC ATC ACC TGC CTG ACC GGA GTC CTG GGT GTG GGC CTG GGT GTG GAG ATC AGC CGC	1035
R L R H S N P R A D P L V C A T G L L G	365
CGG CTC CGC CAC TCC AAC CCC CGG GCT GAT CCC CTG GTC TGT GCC ACT GGC CTC CTG GGC	1095
S A P F L F L S L A C A R G S I V A T Y	385
TCT GCA CCC TTC CTC TTC CTG TCC CTT GCC TGC GCC CGT GGT AGC ATC GTG GCC ACT TAT	1155
I F I F I G E T L L S M N W A I V A D I	405
ATT TTC ATC TTC ATT GGA GAG ACC CTC CTG TCC ATG AAC TGG GCC ATC GTG GCC GAC ATT	1215
L L Y V V I P T R R S T A E A F Q I V L	425
CTG CTG TAC GTG GTG ATC CCT ACC CGA CGC TCC ACC GCC GAG GCC TTC CAG ATC GTG CTG	1275
S H L L G D A G S P Y L I G L I S D R L	445
TCC CAC CTG CTG GGT GAT GCT GGG AGC CCC TAC CTC ATT GGC CTG ATC TCT GAC CGC CTG	1335
R R N W P P S F L S E F R A L Q F S L M	465
CGC CGG AAC TGG CCC CCC TCC TTC TTG TCC GAG TTC CGG GCT CTG CAG TTC TCG CTC ATG	1395
L C A F V G A L G G A A F L G T A I F I	485
CTC TGC GCG TTT GTT GGG GCA CTG GGC GGC GCA GCC TTC CTG GGC ACC GCC ATC TTC ATT	1455
E A D R R R A Q L H V Q G L L H E A G S	505
GAG GCC GAC CGC CGG CGG GCA CAG CTG CAC GTG CAG GGC CTG CTG CAG GAA GCA GGG TCC	1515
T D D R I V V P Q R G R S T R V P V A S	525
ACA GAC GAC CGG ATT GTG GTG CCC CAG CGG GGC CGC TCC ACC CGC GTG CCC GTG GCC AGT	1575
V L I *	529
GTG CTC ATC TGA	1587

GAGGCTGCCGCTCACCTACCTGCACATCTGCCACAGCTGGCCCTGGGCCACCCACGAAGGGCCTGGGCCTAACCCCT
 TGGCCTGGCCCAGCTTCCAGAGGGACCCCTGGGCCGTGTGCCAGCTCCAGACACTACATGGGTAGCTCAGGGGAGGAGG
 TGGGGGTCCAGGAGGGGGATCCCTCTCCACAGGGGCAGCCCCAAGGGCTCGGTGCTATTGTAAACGGAATAAAATTTGT
 AGCCAGAAAAAAGGGCGGCCGC

FIGURE 8B

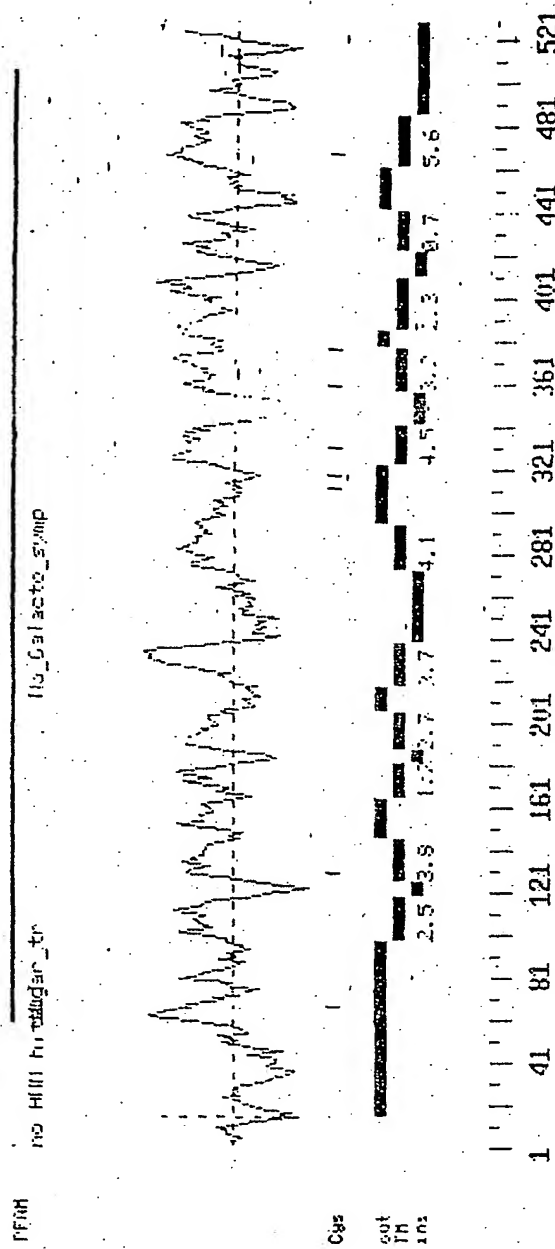


FIGURE 9

Searching for complete domains in PFAM

hmmpfam search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University, School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam

Sequence file: /prod/ddm/wspace/orfanal/oa-script.9015.seq

Query: 46455

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
sugar_tr 1	Sugar (and other) transporter	-63.4	0.00016	
Na_Galacto_symp 1	Sodium:galactoside symporter family	-121.2	0.17	
MCT 1	Monocarboxylate transporter	-208.2	0.32	

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
MCT	1/1	60	473	1	611	-208.2	0.32
sugar_tr	1/1	58	487	1	488	-63.4	0.00016
Na_Galacto_symp	1/1	212	505	1	285	-121.2	0.17

Alignments of top-scoring domains:

MCT: domain 1 of 1, from 60 to 473: score -208.2, E = 0.32

```

-->kpPDGGwGWvVffasFlingfvdGfiksfgvffsellqeetlfnesk
      ++V      in  +   +++++ ++ ++   e++fn+++
46455  60  -----LIVAVLCYINLLNYMDRFTVAGVLPDI---EQFFNIGD 94

      sdvdtAwIgSimlavllfsGPlsSilvnrfGcRivmiaGgllagaGllla
      s  ++I  +++++ +  ++P+ + l +r+  ++ m+ G ++ ++  l +
46455  95  SS--SGLIQTVFISSYMVLAPVFGYLGDRYNRKYLMCGGIAFWSLVTLS 142

      sFst..niwelyltfGvitGlGfgfifqPai.vilgqYFe.....KrRsl
      sF +++++ w+l lt G ++G+G + ++  + ++ ++  F  +++++ +s+
46455  143 SFIPgeHFWLLLLTRG-LVGVGEA-SYSTIApTLIADLFVadqrsRMLSI 190

      AtGiAvaGsGvGtvvfppllqflidnyGsDWrgallilggillncvicGa
      +  GsG+G      +++ ++ d  G DW +al+++ g+  + v++ +
46455  191 FYFAIPVGSGLGY----IAGSKVKDMAG-DWHWALRVTPGLGVVAVLLLF 235

      lllRPlepsvpqdekdkqeetlkeakkkkendtettkeetepkkslpkas
      l +R+      +p++ ++
46455  236 LVVRE---PPRGAVER----- 248

      ilkledakaersvdsLlsskSvgerdksqlsekqksqasgrpsssatavq
46455  - -----

      lvllrsrlekadlplkrvrsvrrrvlSkVSaeSgtdgersSgylNrkdvF
      +s+l
46455  249 ----HSDL----- 252

      YtGsisNvae fkdpdkYrsslhgtrttvgnaesqstlrlddsresgdg
46455  - -----

```

FIGURE 10A

```

dsssedlsektrgdggkkessskeiretikllDfsvlk.nrtFll....
      **** ++ D++ l +n++F+l++ +
46455 253 -----PPLNPTSWWADLRALArNPSFVLsslg 279

.yaisnlfaslGffvPlvfLvsvaikslg.....ldekeAsflls
      a++++ sl++ +P++ L's++ lg++++ +++++ s++
46455 280 fTAVAFVTGSLALWAPAFLLRSRV--VLGetppclpgdsCSSSD-SLIFG 326

      iiGvsnivGRpifGlvADkkgvrpTarhiyifnlsllal..Glttlacp
      i + v +++G+ + +++ r + ++ ll ++l++ ++
46455 327 LITCLTGVLGVGLGVEISRRLRHSNPRADPLVCATGLLGSapFLFLSLAC 376

      latsfwgLvvy CilFGfs.iGsygaLtfvvLvdlvgWlek.....fsnA
      +s++ +++ i+ G + +++a+ + +L v + +++ + f+
46455 377 ARGsIVATYIF-IFIGETILSMNWAIVADILLYVVI-PTRrstaefQIV 424

      fGllllfeGvavLvGPPI....aGlvDakttgdYtvaFyfsGilllllg
      ++ ll G + L+G ++++ + ++++ + ++ ++ ++++
46455 425 LSHLLGDAGSPYLIGLISdrLrRNWPPSF--LSEFRALQFSMLCAFGA 472

      l<--*
      l
46455 473 L 473

sugar_tr: domain 1 of 1, from 58 to 487: score -63.4, E = 0.00016
*->valvaalgGgflfGyDtgvgggflalidflfrfglltssgalaelvg
      al+++ + + + +++++ ++ + f+ + +s+
46455 58 SALIVAVLC-YINLLNYMDRFTVAGVLPDIEQFFNIGDSS----- 96

      ystvlgtglvvsifflGrliGslfaGklgdrfGRKksllialvlfviGall
      +gl+ ++f+ ++ + ++G+lgdr+ Rk+ + +++++ + +l+
46455 97 -----SGLIQTVFISSYMVLAPVFGYLGDRYNRKYLMCGGIAFWSLVTLG 141

      sgaapgytTiGlwafyllivGRvlvGlgvGgasvlvPmYisEiAPkalRG
      s ++pg +f+ll++ R lvG g s ++P++i+ + R
46455 142 SSFIPGE-----HFWLLLLTRGLVGVEASYSTIAPTILIADLFVADQRS 185

      algslyqlaitiGilvAaiiglglnktnndsalsnswgWRiplglqlvpal
      ++ s++ +ai +G +++i g ++ +++d +w R+ gl+ v l
46455 186 RMLSIFYFAIPVGSGLGYIAGSKVKDMAGDW---HWALRVTPGLGVVAVL 232

      llligllflPESPrwLvekgleeArevLaklrgvedvdqelqekaale
      ll++++ P rg + + ++ + +++
46455 233 LLFLVREPP-----RGAVERHSDLPPLNPTSW 260

      atvseekagkaswgelfrgrtrpkvrqrllmgvmlqafqQltGiNaifyY
      + + + l r+++ +l + + +a+ +tG ++ +
46455 261 WA-----DLRALARNPS-----FVLSSLGFTAVAFVTG--SLALW 293

      sptifks.....vGvsdsvasllvtiivgvvnfvfTfvaLiflvD
      +p ++ ++ +++++ +ds +s ++i+g+++ ++ + + l
46455 294 APAFLLRsrvvlggetppCLPGDS-CSSSDSLIFGLITCLTGVLG-VGLGV 341

      rfGRR.....pllllGaagmaicflilgasigvalllllnkpkdpss
      + RR +++++ +pl++ ++ ++ fl+l+ l++ ++
46455 342 EISRRlrhsnpradPLVCATGLLGSAPFLFLS-----LACARGS----- 380

      kaagivaivfillfiafFalgwGpipwilsElFPtkv....Rskalala
      iv++++fi+ + + + w+i++++ v +++Rs+a +
46455 381 -----IVATYIFIFIGE-TLLSMNWAIVADILLYVViPtrRSTAEAFQ 422

      taanwlanfiigflfpyitgaigl....algyvflvfagl...lvlfil
      ++ l + + + py+ g i+ + ++++++ f +l+ +l l+ +
46455 423 IVLSHLLGDAGS---PYLIGLISdrLrrRNWPPSFLSEFRALqfsLMLCAFG 469

```

FIGURE 10B

```

          fvfffvPETkGrtLEeieelf<-*
          ++ ++ .++ G      +i
46455    470 VGALGGAFLG---TAIFIEA      487

Na_Galacto_symp: domain 1 of 1, from 212 to 505: score -121.2, E' = 0.17
          *->qlG.yfffalV...LslagvllwiCf....fgtkEvySssdtreng
          + G+++++ V+++L++++v+ll+++ +++++g E+ sd ++ +
46455    212    MAGdWHWALRVtpgLGVVAVLLLFLVVreppRGAVERH--SDLPPLN 256

          qkttsl1qslkllakNdQ..LliLclaa1fyllainilgg.aqlYYvtYv
          ++      ++l++la+N++ L L++ a + ++ ++l ++a l + v
46455    257    PTSWY-WADLRALARNPSfvLSSLGFTAVAFVTGSLALWapAFLLRSRV 304

          LG.dpelFs.....ylllyn1lvgl1gslLfPrLvkrf..gkktv
          LG++p +++++ +++++ ++l+ +l g g+ L ++r++++ ++
46455    305    LGeTPPCLPgdsccssds1fGLITCLTGVLGVLGVEISRRLrhSNPRA 354

          Fagcivlmvlgsl1iFfvagsslal.ilv1l1flagilqqlvt1lvWv1QV
          + ++lg ++ F++ +++ a + +v++ ++ + + + +W++
46455    355    DPLVCATGLLG-SAPFLFLSLACArgSIVATYIFIFIGETLLSMNWAI-- 401

          IMvsDtVDYGEwktGvRlEGlvvSvflfv1KlGlAlsGalvGwil..gyi
          v+D+ Y t +R+++ ++ l l lG A s l+G+i ++
46455    402    --VADILLYVVIPT-RRSTAEAFQIVLSHL-LGDAGSPYLIGLISdrLRR 447

          GYvanasqststalgQlvfilalFalPpal111laafim1rfYkLtek1la
          + ++ s al+ f l l a++ al +a + ++ f+ + ++
46455    448    NWPPSFL-SEFRALQ---FSLMLCAFVGALGGAFLGTAIFIEADRRRAQ 493

          eIveeLekWrtrkrk<-*
          v L+ + +
46455    494    LHVQGLL---HEAGS      505

```

FIGURE 10C

[illegible]

17/79

Input file 54414; Output File 54414.tra
Sequence length 4632

CACGCGTCCGCCCACGCGTCCGCCCACGCGTCCGAGCCCCCTTCAAGCCTTAGCTTCGGGCTCCAAGCCGACCCCCCTC
CCCCCTCCCTGTCCCCCTTCCCCCTTCTCCCATCCCTCTCTCGGCCACAGCGTCTTGTAGTCTCTCCCTCTACTCCGCAA

TATTTTCTTTCTTTCTCCCTCCTCTCCTCCATTGTTGTTTGATGTTTCCCACTCTTTGAGGAAGG	M V D	3
ATG GTT GAT		9
L E S E V P P L P P R Y R F R D L L L G		23
TTG GAG AGC GAA GTG CCC CCT CTG CCT CCC AGG TAC AGG TTT CGA GAT TTG CTG CTA GGG		69
D Q G W Q N D D K V Q V E F Y M N E N T		43
GAC CAA GGA TGG CAA AAC GAC GAC AGG GTA CAA GTT GAA TTC TAT ATG AAT GAA AAT ACA		129
F K E R L K L F F I K N Q R S S L R I R		63
TTT AAA GAA AGA CTA AAA TTA TTT TTC ATA AAA AAC CAG AGA TCA AGT CTA AGG ATA CGC		189
L F N F S L K L L S C L L Y I I R V L L		83
CTG TTC AAT TTT TCT CTC AAA TTA CTA AGC TGC TTA TTA TAC ATA ATC CGA GTA CTA CTA		249
E N P S Q G N E W S H I F W V N R S L P		103
GAA AAC CCT TCA CAA GGA AAT GAA TGG TCT CAT ATC TTT TGG GTG AAC AGA AGT CTA CCT		309
L W G L Q V S V A L I S L F E T I L L G		123
TTG TGG GGC TTA CAG GTT TCA GTG GCA TTG ATA AGT CTG TTT GAA ACA ATA TTA CTT GGT		369
Y L S Y K G N I W E Q I L R I P F I L E		143
TAT CTT AGT TAT AAG GGA AAC ATC TGG GAA CAG ATT TTA CGA ATA CCC TTC ATC TTG GAA		429
I I N A V P F I I S I F W P S L R N L F		163
ATA ATT AAT GCA GTT CCC TTC ATT ATC TCA ATA TTC TGG CCT TCC TTA AGG AAT CTA TTT		489
V P V F L N C W L A K H A L D N M I N D		183
GTC CCA GTC TTT CTG AAC TGT TGG CTT GCC AAA CAT GCC TTG GAT AAT ATG ATT AAT GAT		549
L H R A I Q R T Q S A M F N Q V L I L I		203
CTA CAC AGA GCC ATT CAG CGT ACA CAG TCT GCA ATG TTT AAT CAA GTT TTG ATT TTA ATA		609
S T L L C L I F T C I C G I Q H L E R I		223
TCT ACA TTA CTA TGC CTT ATC TTC ACC TGC ATT TGT GGG ATC CAA CAT CTG GAA CGA ATA		669
G K R L N L F D S L Y F C I V T F S T V		243
GGA AAG AGG CTG AAT CTC TTT GAC TCC CTT TAT TTC TGC ATT GTG ACG TTT TCT ACT GTG		729
G F G D V T P E T W S S K L F V V A M I		263
GGC TTC GGG GAT GTC ACT CCT GAA ACA TGG TCC TCC AAG CTT TTT GTA GTT GCT ATG ATT		789
C V A L V V L P I Q F E Q L A Y L W M E		283
TGT GTT GCT CTT GTG GTT CTA CCC ATA CAG TTT GAA CAG CTG GCT TAT TTG TGG ATG GAG		849
R Q K S G G N Y S R H R A Q T E K H V V		303
AGA CAA AAG TCA GGA GGA AAC TAT AGT CGA CAT AGA GCT CAA ACT GAA AAG CAT GTC GTC		909
L C V S S L K I D L L M D F L N E F Y A		323
CTG TGT GTC AGC TCA CTG AAG ATT GAT TTA CTT ATG GAT TTT TTA AAT GAA TTC TAT GCT		969

FIGURE 12A

H P R L Q D Y Y V V I L C P T E M D V Q 343
CAT CCT AGG CTC CAG GAT TAT TAT GTG GTG ATT TTG TGT CCT ACT GAA ATG GAT GTA CAG 1029

V R R V L Q I P M W S Q R V I Y L Q G S 363
GTT CGA AGG GTA CTG CAG ATT CCA ATG TGG TCC CAA CGA GTT ATC TAC CTT CAA GGT TCA 1089

A L K D Q D L L R A K M D D A E A C F I 383
GCC CTT AAA GAT CAA GAC CTA TTG AGA GCA AAG ATG GAT GAC GCT GAG GCC TGT TTT ATT 1149

L S S R C E V D R T S S D H Q T I L R A 403
CTC AGT AGC CGT TGT GAA GTG GAT AGG ACA TCA TCT GAT CAC CAA ACA ATT TTG AGA GCA 1209

W A V K D F A P N C P L Y V Q I L K P E 423
TGG GCT GTG AAA GAT TTT GCT CCA AAT TGT CCT TTG TAT GTC CAG ATA TTA AAG CCT GAA 1269

N K F H I K F A D H V V C E E E F K Y A 443
AAT AAA TTT CAC ATC AAA TTT GCT GAT CAT GTT GTT TGT GAA GAA GAG TTT AAA TAC GCC 1329

M L A L N C I C P A T S T L I T L L V H 463
ATG TTA GCT TTA AAC TGT ATA TGC CCA GCA ACA TCT ACA CTT ATT ACA CTA CTG GTT CAT 1389

T S R G Q E G Q Q S P E Q W Q K M Y G R 483
ACC TCT AGA GGG CAA GAA GGC CAG CAA TCG CCA GAA CAA TGG CAG AAG ATG TAC GGT AGA 1449

C S G N E V Y H I V L E E S T F F A E Y 503
TGC TCC GGG AAT GAA GTC TAC CAC ATT GTT TTG GAA GAA AGT ACA TTT TTT GCT GAA TAT 1509

E G K S F T Y A S F H A H K K F G V C L 523
GAA GGA AAG AGT TTT ACA TAT GCC TCT TTC CAT GCA CAC AAA AAG TTT GGC GTC TGC TTG 1569

I G V R R E D N K N I L L N P G P R Y I 543
ATT GGT GTT AGG AGG GAG GAT AAT AAA AAC ATT TTG CTG AAT CCA GGT CCT CGA TAC ATT 1629

M N S T D I C F Y I N I T K E E N S A F 563
ATG AAT TCT ACG GAC ATA TGC TTT TAT ATT AAT ATT ACC AAA GAA GAG AAT TCA GCA TTT 1689

K N Q D Q Q R K S N V S R S F Y H G P S 583
AAA AAC CAA GAC CAG CAG AGA AAA AGC AAT GTG TCC AGG TCG TTT TAT CAT GGA CCT TCC 1749

R L P V H S I I A S M G T V A I D L Q D 603
AGA TTA CCT GTA CAT AGC ATA ATT GCC AGC ATG GGT ACT GTG GCT ATA GAC CTG CAA GAT 1809

T S C R S A S G P T L S L P T E G S K E 623
ACA AGC TGT AGA TCA GCA AGT GGC CCT ACC CTG TCT CTT CCT ACA GAG GGA AGC AAA GAA 1869

I R R P S I A P V L E V A D T S S I Q T 643
ATA AGA AGA CCT AGC ATT GCT CCT GTT TTA GAG GTT GCA GAT ACA TCA TCG ATT CAA ACA 1929

C D L L S D Q S E D E T T P D E E M S S 663
TGT GAT CTT CTA AGT GAC CAA TCA GAA GAT GAA ACT ACA CCA GAT GAA GAA ATG TCT TCA 1989

N L E Y A K G Y P P Y S P Y I G S S P T 683
AAC TTA GAG TAT GCT AAA GGT TAC CCA CCT TAT TCT CCA TAT ATA GGA AGT TCA CCC ACT 2049

F C H L L H E K V P F C C L R L D K S C 703
TTT TGT CAT CTC CTT CAT GAA AAA GTA CCA TTT TGC TGC TTA AGA TTA GAC AAG AGT TGC 2109

FIGURE 12B

Q	H	N	Y	Y	E	D	A	K	A	Y	G	F	K	N	K	L	I	I	V	723
CAA	CAT	AAC	TAC	TAT	GAG	GAT	GCA	AAA	GCC	TAT	GGA	TTT	AAA	AAT	AAA	CTA	ATT	ATA	GTT	2169
A	A	E	T	A	G	N	G	L	Y	N	F	I	V	P	L	R	A	Y	Y	743
GCA	GCT	GAA	ACA	GCT	GGA	AAT	GGA	TTA	TAT	AAC	TTT	ATT	GTT	CCT	CTC	AGG	GCA	TAT	TAT	2229
R	P	K	K	E	L	N	P	I	V	L	L	L	D	N	P	L	D	D	L	763
AGA	CCA	AAG	AAA	GAA	CTT	AAT	CCC	ATA	GTA	CTG	CTA	TTG	GAT	AAC	CCC	CTA	GAT	GAC	TTA	2289
L	R	C	G	V	T	F	A	A	N	M	V	V	V	D	K	E	S	T	M	783
CTC	AGG	TGT	GGA	GTG	ACT	TTT	GCT	GCT	AAT	ATG	GTG	GTT	GTG	GAT	AAA	GAG	AGC	ACC	ATG	2349
S	A	E	E	D	Y	M	A	D	A	K	T	I	V	N	V	Q	T	L	F	803
AGT	GCC	GAG	GAA	GAC	TAC	ATG	GCA	GAT	GCC	AAA	ACC	ATT	GTG	AAC	GTG	CAG	ACA	CTC	TTC	2409
R	L	F	S	S	L	S	I	I	T	E	L	T	H	P	A	N	M	R	F	823
AGG	TTG	TTT	TCC	AGT	CTC	AGT	ATT	ATC	ACA	GAG	CTA	ACT	CAC	CCC	GCC	AAC	ATG	AGA	TTC	2469
M	Q	F	R	A	K	D	C	Y	S	L	A	L	S	K	L	E	K	K	E	843
ATG	CAA	TTC	AGA	GCC	AAA	GAC	TGT	TAC	TCT	CTT	GCT	CTT	TCA	AAA	CTG	GAA	AAG	AAA	GAA	2529
R	E	R	G	S	N	L	A	F	M	F	R	L	P	F	A	A	G	R	V	863
CGG	GAG	AGA	GGC	TCT	AAC	TTG	GCC	TTT	ATG	TTT	CGA	CTG	CCT	TTT	GCT	GCT	GGG	AGG	GTG	2589
F	S	I	S	M	L	D	T	L	L	Y	Q	S	F	V	K	D	Y	M	I	883
TTT	AGC	ATC	AGT	ATG	TTG	GAC	ACT	CTG	CTG	TAT	CAG	TCA	TTT	GTG	AAG	GAT	TAT	ATG	ATT	2649
S	I	T	R	L	L	L	G	L	D	T	T	P	G	S	G	F	L	C	S	903
TCT	ATC	ACG	AGA	CTT	CTG	TTG	GGA	CTG	GAC	ACT	ACA	CCA	GGA	TCT	GGG	TTT	CTT	TGT	TCT	2709
M	K	I	T	A	D	D	L	W	I	R	T	Y	A	R	L	Y	Q	K	L	923
ATG	AAA	ATC	ACT	GCA	GAT	GAC	TTA	TGG	ATC	AGA	ACT	TAT	GCC	AGA	CTT	TAT	CAG	AAG	TTG	2769
C	S	S	T	G	D	V	P	I	G	I	Y	R	T	E	S	Q	K	L	T	943
TGT	TCT	TCT	ACT	GGA	GAT	GTT	CCC	ATT	GGA	ATC	TAC	AGG	ACT	GAG	TCT	CAG	AAA	CTT	ACT	2829
T	S	E	S	R	K	I	A	S	Q	S	Q	I	S	I	S	V	E	E	W	963
ACA	TCT	GAG	TCT	CGA	AAA	ATA	GCA	TCA	CAA	TCT	CAA	ATA	TCT	ATC	AGT	GTA	GAA	GAG	TGG	2889
E	D	T	K	D	S	K	E	Q	G	H	H	R	S	N	H	R	N	S	T	983
GAA	GAC	ACC	AAA	GAC	TCC	AAA	GAA	CAA	GGG	CAC	CAC	CGC	AGC	AAC	CAC	CGC	AAC	TCA	ACA	2949
S	S	D	Q	S	D	H	P	L	L	R	R	K	S	M	Q	W	A	R	R	1003
TCC	AGT	GAC	CAG	TCG	GAC	CAT	CCC	TTG	CTG	CGG	AGA	AAA	AGC	ATG	CAG	TGG	GCC	CGA	AGA	3009
L	S	R	K	G	P	K	H	S	G	K	T	A	E	K	I	T	Q	Q	R	1023
CTG	AGC	AGA	AAA	GGC	CCA	AAA	CAC	TCT	GGT	AAA	ACA	GCT	GAA	AAA	ATA	ACC	CAG	CAG	CGA	3069
L	N	L	Y	R	R	S	E	R	Q	E	L	A	E	L	V	K	N	R	M	1043
CTG	AAC	CTC	TAC	AGG	AGG	TCA	GAA	AGA	CAA	GAG	CTT	GCT	GAA	CTT	GTG	AAA	AAT	AGA	ATG	3129
K	H	L	G	L	S	T	V	G	Y	D	E	M	N	D	H	Q	S	T	L	1063
AAA	CAC	TTG	GGT	CTT	TCT	ACA	GTG	GGA	TAT	GAT	GAA	ATG	AAT	GAT	CAT	CAA	AGT	ACC	CTC	3189
S	Y	I	L	I	N	P	S	P	D	T	R	I	E	L	N	D	V	V	Y	1083
TCC	TAC	ATC	CTG	ATT	AAC	CCA	TCT	CCA	GAT	ACC	AGA	ATA	GAG	CTG	AAT	GAT	GTT	GTA	TAC	3249

FIGURE 12C

L I R P D P L A Y L P N S E P S R R N S 1103
TTA ATT CGA CCA GAT CCA CTG GCC TAC CTG CCA AAC AGT GAG CCC AGT CGA AGA AAC AGC 3309
I C N V T G Q D S R E E T Q L * 1119
ATC TGC AAT GTC ACT GGT CAA GAT TCT CGG GAG GAA ACT CAA CTT TGA 3357
TAAAAATAAAATGAGAACTTTTTTCTACAAAGACCTTGCTTGAAACCACAAAAGTTTTGCTGGCAGAAAGAACTA
GATGGAAATATATGTAATTCTCTCATATTTAAAAACGTAATCTCTTCTTTAGAAAGTATAGATCATTGAAACTTAAT
GTACTACTTACTGGTACTCTCCCTATTAATATTTGAAGGACCTCAATGGAATAAATTTGAAAAGCTAAATTAAAAATACA
AAAATTTAAATCTGACATTTAATTGTTTTATAATAATCCAACTCTATGAAAGCAATTTTAAAAATTATTAAGGTTTTTA
TGAAGTTGACAAAATCTAACTATATTTGGTGCATCACAATGGACACAGAATGCTGCTGCTCCTCTTAAAAATTAAATGT
GTCATATTATATTCTTTAACTTACTGTTTTACAAAATTGAGCTCATCGTAAATGTCTAGTCTTCTCACATAGAGATTA
ACCAACAACTTGTGTGGCTGACTTTTTGTGTAAGAATCATAGTTTGCTTTAGAATACAAATCTTTAAGTCATTTTAACT
TTTTTTCTGCCTTACGATATAAAAAATATTTATCTTAGAATTTGAGATGTTTCATAGCATGTTTTATTACATTGAAGAAA
CTAAACATAAATGAAAAGAAACACTAGGTTCTGCACTTTTTGGTAACTTTATGTCTAGCAAATATTTTATGCCAAGA
AAAGCATACTATAAAGCAAATATCTATTATTCTCTAAACGAATGCCTAGCATAGAGAAAATACTTAATACACATTTGT
TGACTTAAATTTAATTCAAGGATTGAAAAATTAAGTGGATATCTTGAAATATACAGTAATGATTGTCCTTAGACTCTTG
AACTTTACCATCTTTCCTATTCATATATCTATATAGTAAATTTCACTAGAAAAATCTTTTAAAAATTGACAGAAGATAA
TTTATACCTTTTATGGACTCTGAAGACACTTCAAAACATTAAAAGTCCTTATGTCTTTGGTAATGAAACATACACTCAA
TGANGATGTATTAAATTTTGACTT

FIGURE 12D

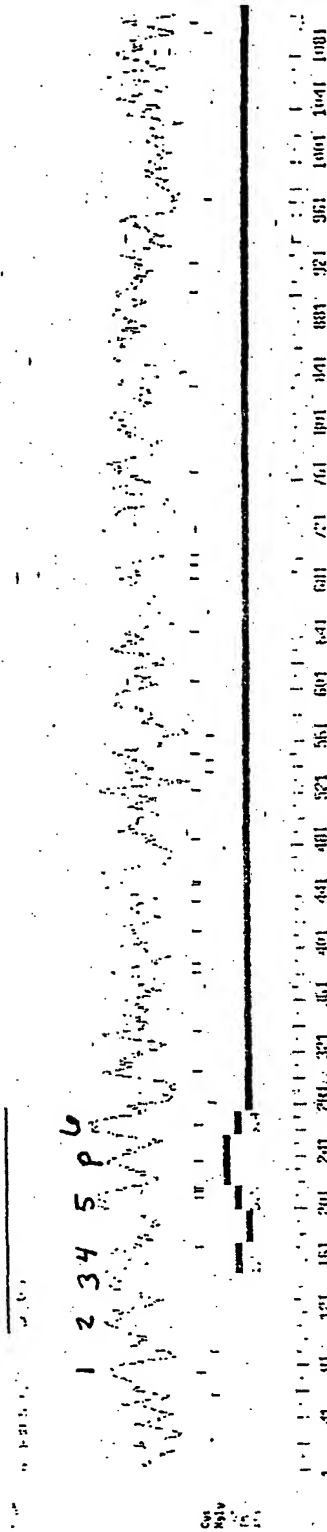


FIGURE 13

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL)

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.3743.seq

Query: Fbh54414

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
ion_trans	Ion transport protein	62.4	9.9e-15	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
ion_trans	1/1	104	277	1	223	62.4	9.9e-15

Alignments of top-scoring domains:

ion_trans: domain 1 of 1, from 104 to 277: score 62.4, E = 9.9e-15

```

      *->ilfildllfvllfllleivlkfiayglkstsniaakylksifnildll
      +++++ ++ + ++ +i+l + +y ++ ++ +++++i +il ++
Fbh54414 104 LWGLQVSVALISLFETILGYLSYKGN-----IWEQILRIPFILEII 145

      ailplllllvflsgteqvakkrlrerfslelsqwyrrilrflrllrllr
      ++p+++++++ ++ 1 + ++L ++
Fbh54414 146 NAVPFIISIFWPSLRN-----LFVPVFLNCW- 171

      lLrllrllrrletlfefelgtlaWslqlsgralksilrflllllllligf
      1 + + +++ + +l + ++r++ +++++ +l+l+ 1++
Fbh54414 172 ---LAKHALENM-----INDL---HRAIQRTQSAMFNQVLILISTLLCL 209

      svigyllfkgyedlsenevgnsefssyfdafyflfvltttvGfGdlvpv
      ++ + +++e+ ++ ++fd++yf++vt++tvGfGd++p+
Fbh54414 210 IFTCICGIIQHLE-----IGKRLNLFDSLYFCIVTFSTVGFQDVTPE 251

      .wlgiiiffvlfliivgllllnllliavi<-*
      +w++++f+ + +i+v+l++l + + +
Fbh54414 252 tWSSKLFV-VAMICVALVVLPIQFEQL 277

```

FIGURE 14

CLUSTAL W (1.74) multiple sequence alignment

```

54414.prot -----MVDLESEVPPLPPR---YRFRDL--LLGDEGWQN
AF089730 MARAKLPRSPSEKAGPGDTPAGSAAPEEPHGLSPLLPTRGGSVGSVDVGQRLHVEDFSL

54414.prot DD---RVQVEFYMNENTFKERLKLFFIKQRSSLRIHLNFNFKLLSCLLYIIRVLIENP
AF089730 DSSLQVQVEFYMNENTFKERLKLFFIKQRSSLRIHLNFNFKLLTCLLYIVRVLIENP

54414.prot SQGN-----EWSHIFWVNRSLHNLGLQVSVALISLFETILLGYLSY
AF089730 DQGIGCWGCTKYNITFNGSSSEFWAPILWVERKMAWVQIVATISFLEMLLIYLSY

54414.prot KGNIEWQLIRIPFILEIINAVFFIISLFWPSLNLFFVFLNCLWLAHGALENMINDLHRA
AF089730 KGNIEWQLIFHVSFVLEMINTLFFIITVFWPPLNLFILVFLNCLWLAHGALENMINDFHRA

54414.prot IQRFOSAMFNQVLILISTLLCLFTCTCGIHLERIGKRLALFDSL YPCIVTFSTVGFPGD
AF089730 ILRFOSAMFNQVLILFCTLLCLVFTGTGCIHLERAGGNLNLTSFYFCIVTFSTVGFPGD

54414.prot VTFETWSSKLFVVAMICVALVVLPIQEPOLAYLWMERQKSGGNYSRHRAQTEKHVVLVCVS
AF089730 VTFETWSSKLFVVAMICVALVVLPIQEPOLAYLWMERQKSGGNYSRHRAQTEKHVVLVCVS

54414.prot SLKIDLLMDFLNEFYAHPRLQDYVVILCPTMDVQVRRVLQIPMWSQRVIYLOGSALKD
AF089730 SLKIDLLMDFLNEFYAHPRLQDYVVILCPTMDVQVRRVLQIPMWSQRVIYLOGSALKD

54414.prot QDLLRAKMDDAEACFILSSRCEVDRTSSDHQITLRAWAVKDFAPNCPLYVQILKPKENKFH
AF089730 QDLIRAKMDNGEACFILSSRNEVDRTAADHQITLRAWAVKDFAPNCPLYVQILKPKENKFH

54414.prot IKFADHVVCEEFFKYAMLALNCICPATSTLITLLVHTSRQEGQSQSPQWQMYGRCSGN
AF089730 VKFADHVVCEEFFKYAMLALNCICPATSTLITLLVHTSRQEGQSQSPQWQMYGRCSGN

54414.prot EVVHIVLEESTFFAEYEGKSFTYASFHAHKKFGVCLIGVRREDNIGNILLNPGPRYIMNST
AF089730 EVVHIRMGDSKFFREYEGKSFTYAAFHAHKKYGVCILGLKREENKSILLNPGPRHILAAS

54414.prot DICFYINTITKEENSA--FKNQDQQRKSNVS--SFYHGPSRLPVHSIIASMGTVAIDLQDT
AF089730 DTCFYINTITKEENSAFIFKQEEKQNRRLAGQALYEGPSRLPVHSIIASM--VAMDQNT

54414.prot SCRASAGPT-----LSLPTEGSKEIRRPISAPVLEVADTSSIQTCDLLSDQSEDETP-
AF089730 DCRPSQGGSGGGGKLTLPTEGSGSRRPISAPVLELADSSALLPCDLLSDQSEDETPS

54414.prot DEEMSSNLEYAKGYPPYSPYIGSSPTFCHLLHEKVFPCLRLDKSCQHNYVEDAKAYGFK
AF089730 DDEGLSVVEYVKGYPNSPYIGSSPTLCHLLPVKAPFCLRLDKGCHNSYEDAKAYGFK

54414.prot NKLIIVAAETAGNGLYNFIVPLRAYYRPKKELNPVLLIDN-P-----
AF089730 NKLIIVSAETAGNGLYNFIVPLRAYYRSRELNPVLLIDNKPDHFLAICCFPMVYYM

54414.prot -----LDLLRCGVTFANMNVVDKESTMSAEEDYMADAKTIVNVQTLFRLFSLSIIT
AF089730 EGSVDNLDLSLQCGIYADNLVVVDKESTMSAEEDYMADAKTIVNVQTMFRLFPFSLSIIT

```

G-hydrophobic G
potassium channel
signature domain

FIGURE 15A

54414.prot
AF089730
ELTHPANMRFMQFRAKDCYSLALS KLEKKEFEREGSNLAFMRLPFAAGRVFSISM LDTLL
ELTHPSNMRFMQFRAKDSYSLALS KLEKQERENGSNLAFMRLPFAAGRVFSISM LDTLL
.....

54414.prot
AF089730
YQSFVKDYMISITRLLGLDTPGSGFLCSMKITADDLWIRTYARLYOKLCSSTGDVPIG
YQSFVKDYMISITRLLGLDTPGSGFLCAMKVTEDDLWIRTYGRLFOKLCSSSAEIPIG
.....

54414.prot
AF089730
IYRTESQKLTTSES RKIASQSQISISVEEWEDTKDSK-----EQGHRSNHRNSTSSD
IYRTECH-VFSSEPHDLRAQSQISVN MEDCEDTREAKGPWGTRAASGGGSTHGRHGSAD

54414.prot
AF089730
QSDHPLLRKSMQWARRLSFKGPKHSGK---TAEKITQORLNLYRRSERQFLAELVKNRN
PVEHPLLRKSLQWARKLSRKSSKQAGKAPMTTDWITQORLSLYRRSERQELSELVKNRN

54414.prot
AF089730
KHLGLSTVGY-----DEMND-HQSTLSYILINPSDTRI ELDV
KHLGLPTTGIEDVANLTASDVMNRVNLGYLQDEMNDHHQNTLSYVLINPPDTRLEPNDI

54414.prot
AF089730
VYLIRPDPLAYLPNSEPSRRNSICNVTG---QDSREETQL
VYLIRSDPLAHVTSSSQSRKSSCSNKLSSCNPETRDETQL

FIGURE 15B

Input file Fbh53763pat.seq; Output File Fbh53763pat.tta
Sequence length 2847

```

CCACGCGTCCGGCCCTGCTTCGGATGGCGGCGGAGGTTGATGGCGAGTGGTGCTGAAGGGACAGCTCCAGCAGTGG
CTGATTTGGGGGAGAAACAAAATCTGCAGATGGAATCCGAGCAGGGCGACTTCACCTTCAAGTGGTGAGCTCTCCTGAC
CTGCGGCCAGTCTCCACTCCATTACGGCCAGCCGATCTGCCCGCTCCCGAGGGGTGCGGCAGTGCCGGCTGGACCCG
CCCCGAGCTCCATGTTTGGCCAAACCTGCGCGATGGTGACTGTGGGCGCGGAGGTTGGCGACTGGCAAAATCCGCAGAT
CACAGAATGAAGSGCGGGAGCGCGCGCGCGCGCGCGGGGGCTTCTCCCCCACCCAGCGCCAGGGAAGCGGCTCA
ACCACCTGAATCCGGAAGACGCCAACAAAGTAGTTTCTCGTCCGAGAAGGGCGGCTCACCTGGGCGCCAAGACTCAGTCC
CGCTGCCCAGAGAACCTCGTCCACTCGGAAACCAAGCAGAACCCTTTCTCTCGGTCTCGTTAAGTCATGTCTGAGT

      M   G   K   I   E   N   N   E   R   V   I   L   N   V   G   G   T   R
CACAGAG ATG GGC AAG ATC GAG AAC AAC GAG AGG GTG ATC CTC AAT GTC GGG GGC ACC CGG 18
      54
      H   E   T   Y   R   S   T   L   K   T   L   P   G   T   R   L   A   L   L   A
CAC GAA ACC TAC CGC AGC ACC CTC AAG ACC CTG CCT GGA ACA CGC CTG GCC CTT CTT GCC 38
      114
      S   S   E   P   P   G   D   C   L   T   T   A   G   D   K   L   Q   P   S   P
TCC TCC GAG CCC CCA GGC GAC TGC TTG ACC ACG GCG GGC GAC AAG CTG CAG CCG TCG CCG 58
      174
      P   P   L   S   P   P   P   R   A   P   P   L   S   P   G   P   G   G   C   F
CCT CCA CTG TCG CCG CCG CCG AGA GCG CCC CCG CTG TCC CCC GGG CCA GGC GGC TGC TTC 78
      234
      E   G   G   A   G   N   C   S   S   R   G   G   R   A   S   D   H   P   G   G
GAG GGC GGC GCG GGC AAC TGC AGT TCC CGC GGC GGC AGG GCC AGC GAC CAT CCC GGT GGC 98
      294
      G   R   E   F   F   F   D   R   H   P   G   V   F   A   Y   V   L   N   Y   Y
GGC CGC GAG TTC TTC TTC GAC CGG CAC CCG GGC GTC TTC GCC TAT GTG CTC AAT TAC TAC 118
      354
      R   T   G   K   L   H   C   P   A   D   V   C   G   P   L   F   E   E   E   L
CGC ACC GGC AAG CTG CAC TGC CCC GCA GAC GTG TGC GGG CCG CTC TTC GAG GAG GAG CTG 138
      414
      A   F   W   G   I   D   E   T   D   V   E   P   C   C   W   M   T   Y   R   Q
GCC TTC TGG GGC ATC GAC GAG ACC GAC GTG GAG CCC TGC TGC TGG ATG ACC TAC CGG CAG 158
      474
      H   R   D   A   E   E   A   L   D   I   F   E   T   P   D   L   I   G   G   D
CAC CGC GAC GCC GAG GAG GCG CTG GAC ATC TTC GAG ACC CCC GAC CTC ATT GGC GGC GAC 178
      534
      P   G   D   D   E   D   L   A   A   K   R   L   G   I   E   D   A   A   G   L
CCC GGC GAC GAC GAG GAC CTG GCG GCC AAG AGG CTG GGC ATC GAG GAC GCG GCG GGG CTC 198
      594
      G   G   P   D   G   K   S   G   R   W   R   R   L   Q   P   R   M   W   A   L
GGG GGC CCC GAC GGC AAA TCT GGC CGC TGG AGG AGG CTG CAG CCC CGC ATG TGG GCC CTC 218
      654
      F   E   D   P   Y   S   S   R   A   A   R   F   I   A   F   A   S   L   F   F
TTC GAA GAC CCC TAC TCG TCC AGA GCC GCC AGG TTT ATT GCT TTT GCT TCT TTA TTC TTC 238
      714
      I   L   V   S   I   T   T   F   C   L   E   T   H   E   A   F   N   I   V   K
ATC CTG GTT TCA ATT ACA ACT TTT TGC CTG GAA ACA CAT GAA GCT TTC AAT ATT GTT AAA 258
      774
      N   K   T   E   P   V   I   N   G   T   S   V   V   L   Q   Y   E   I   E   T
AAC AAG ACA GAA CCA GTC ATC AAT GGC ACA AGT GTT GTT CTA CAG TAT GAA ATT GAA ACG 278
      834

```

FIGURE 16A

D" P A L T Y V E G V C V V W F T F E F I	298
GAT CCT GCC TTG ACG TAT GTA GAA GGA GTG TGT GTG GTG TGG TTT ACT TTT GAA TTT TTA	894
V R I V F S P N K L E F I K N L L N I I	318
GTC CGT ATT GTT TTT TCA CCC AAT AAA CTT GAA TTC ATC AAA AAT CTC TTG AAT ATC ATT	954
D F V A I L P F Y L E V G L S G L S S K	338
GAC TTT GTG GCC ATC CTA CCT TTC TAC TTA GAG GTG GGA CTC AGT GGG CTG TCA TCC AAA	1014
A A K D V L G F L R V V R F V R I L R I	358
GCT GCT AAA GAT GTG CTT GGC TTC CTC AGG GTG GTA AGG TTT GTG AGG ATC CTG AGA ATT	1074
F K L T R H F V G L R V L G H T L R A S	378
TTC AAG CTC ACC CGC CAT TTT GTA GGT CTG AGG GTG CTT GGA CAT ACT CTT CGA GCT AGT	1134
T N E F L L L I I F L A L G V L I F A T	398
ACT AAT GAA TTT TTG CTG CTG ATA ATT TTC CTG GCT CTA GGA GTT TTG ATA TTT GCT ACC	1194
M I Y Y A E R V G A Q P N D P S A S E H	418
ATG ATC TAC TAT GCC GAG AGA GTG GGA GCT CAA CCT AAC GAC CCT TCA GCT AGT GAG CAC	1254
T Q F K N I P I G F W W A V V T M T T L	438
ACA CAG TTC AAA AAC ATT CCC ATT GGG TTC TGG TGG GCT GTA GTG ACC ATG ACT ACC CTG	1314
G Y G D M Y P Q T W S G M L V G A L C A	458
GGT TAT GGG GAT ATG TAC CCC CAA ACA TGG TCA GGC ATG CTG GTG GGA GCC CTG TGT GCT	1374
L A G V L T I A M P V P V I V N N F G M	478
CTG GCT GGA GTG CTG ACA ATA GCC ATG CCA GTG CCT GTC ATT GTC AAT AAT TTT GGA ATG	1434
Y Y S L A M A K Q K L P R K R K K H I P	498
TAC TAC TCC TTG GCA ATG GCA AAG CAG AAA CTT CCA AGG AAA AGA AAG AAG CAC ATC CCT	1494
P A P Q A S S P T F C K T E L N M A C N	518
CCT GCT CCT CAG GCA AGC TCA CCT ACT TTT TGC AAG ACA GAA TTA AAT ATG GCC TGC AAT	1554
S T Q S D T C L G K D N R L L E H N R S	538
AGT ACA CAG AGT GAC ACA TGT CTG GGC AAA GAC AAT CGA CTT CTG GAA CAT AAC AGA TCA	1614
V L S G D D S T G S E P P L S P P E R L	558
GTG TTA TCA GGT GAC GAC AGT ACA GGA AGT GAG CCG CCA CTA TCA CCC CCA GAA AGG CTC	1674
P I R R S S T R D K N R R G E T C F L L	578
CCC ATC AGA CGC TCT AGT ACC AGA GAC AAA AAC AGA AGA GGG GAA ACA TGT TTC CTA CTG	1734
T T G D Y T C A S D G G I R K G Y E K S	598
ACG ACA GGT GAT TAC ACG TGT GCT TCT GAT GGA GGG ATC AGG AAA GGA TAT GAA AAA TCC	1794
R S L N N I A G L A G N A L R L S P V T	618
CGA AGC TTA AAC AAC ATA GCG GGC TTG GCA GGC AAT GCT CTG AGG CTC TCT CCA GTA ACA	1854
S P Y N S P C P L R R S R S P I P S I L	638
TCA CCC TAC AAC TCT CCT TGT CCT CTG AGG CGC TCT CGA TCT CCC ATC CCA TCT ATC TTG	1914
* TAA	639
	1917

FIGURE 16B

ACCAAACAACCAAACTGCATCAGTCGGCTAAATGTATTAATTCAGYGCCTGTTACCCATAATGGAATAATTAAAT
GTAGAGTTATCCAGGCTCCATTAATACAGTATAAATCTTGCGTGATACTACAAATTTGAAGTCAGAAATGCCACTTGGG
TAGCTAATGAATCTTACCCAGGCTTTAAAGATTGTCTAAAGTAGTGCTAAGATCCCTCCTATTAATTGCCCTGATATCC
TTTTGCAATAAAATGACAGATAGTGTGATATTGACCAGTGCACTAATATATAAACATACCCTCAGGGAGATATATTT
AAAACAGTGTGCTTCCAAATGCCAACCACCTTCATTGGAACCTTATTTCTTGTGA

FIGURE 16C

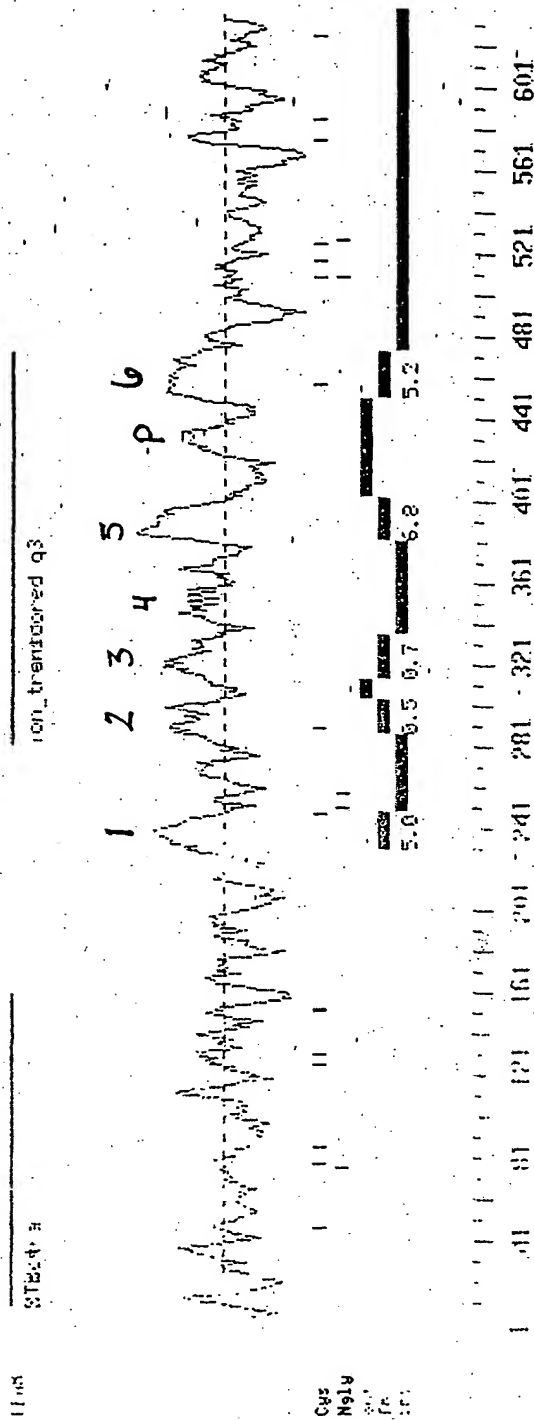


FIGURE 17

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam

Sequence file: /prod/ddm/wspace/orfanal/oa-script.4688.seq

Query: Fbh53763

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
K_tetra	K ⁺ channel tetramerisation domain	156.7	4e-43	
ion_trans	Ion transport protein	116.9	3.9e-31	
oxidored_q3	NADH-ubiquinone/plastoquinone oxidoreduct	-81.7	5.6	

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
K_tetra	1/1	8	156	1	111	156.7	4e-43
oxidored_q3	1/1	317	467	1	177	-81.7	5.6
ion_trans	1/1	281	472	1	223	116.9	3.9e-31

Alignments of top-scoring domains:

K_tetra: domain 1 of 1, from 8 to 156: score 156.7, E = 4e-43

```

*->ErvrLNVGGkrFeTskstLtrfkpdTLgrllktdsd.....
Erv+LNVGG+r+eT++sTB ++ p T+L l++ s++++ ++ +
Fbh53763      8  ERVILNVGGTRHETYSTLKTL-PGTRLALLAS--SEppgdclttag 51

.....vhearlrld
++ +++++ +++++ ++ +++++ ++ +++++ +++++ d
Fbh53763     52 dklqpsppplsppprapplsppgpgcfeeggagncssrggrA-----SD 94

fyddetgEyFFDRsPkhFetILnfYRtGdGkLhrp.evcldsfleEleFy
+ E+FFDR+P++F ++Ln+YRt GkLh+p +vc f+eEl+F+
Fbh53763     95 HPGGGR-EFFDRHPGVFAYVLNYYRT--GKLHCPaDVCGLPFEEELAFW 141

gldelaiesCcedey<-*
g+de ++e+Cc+++Y
Fbh53763    142 GIDETDVEPCWMTY      156

```

oxidored_q3: domain 1 of 1, from 317 to 467: score -81.7, E = 5.6

```

*->mtiyivliLsillvlGflgVaskpsPiYgaLgLivaggvGCGlvlslg
+ +v iL + l +G+ g++sk + + + +v ++ + + + l
Fbh53763    317 IIDFVAILPFYLEVGLSGLSSKAAKDVLGFLRVVRFVRI-LRIFKLT 362

gsFvalvlFLIYLGGMlVVfGytvalateeyPEaWgsnkvvtigdgval
Fv+l ++ g t t e+ + ++i ++l
Fbh53763    363 RHEVGLRVL-----GHTLRASSTNEF-----L---LLI---IFL 389

vlgllievlLvglvl.....gwtevvivvaltgldwviYdvsgs
lg+li++ +++++ ++ + +++++ e + + + + G w + v + +
Fbh53763    390 ALGVLI FATMIYYAervgaqndpSASEHTQFKNIP-IGFW--WAVVTM- 435

liredlsGvaaLYscgvwmfevaGwvLLvalfvvieltR<-*
++ G +Y +w + G L al +v+++++
Fbh53763    436 ----TTLGYGDMYPQ-TWSGMLVG--ALCAGVLTIAM 467

```

FIGURE 18A


```

ion_trans: domain 1 of 1, from 281 to 472: score 116.9, E = 3.9e-31
*->ilfildllflvllflleivlkfiayglkstsniaakylksifnildll
1++++ ++v++f++e+++++ ++k      ++k+ ni+d+
Fbh53763 281 ALTYVEGVCVVWFTFEFLVRIVFSPNK-----LEFIKNLLNIIDFV 321
ailpllllllvflsgteqvakkrlrer.f.slelsq.wyyrilrflrlLr
ailp++l ++l      +++++ ++k      +flr++r
Fbh53763 322 AILPFYLEVGL-----SgLSsKAAKDvL-----GFLRVVR 351
llRlLrllrllrrletlfefelgtlaWslqslgralksilrfl1111111
++R  +lr++ +++      +++      1+ lg++l++ ++ +lll++l
Fbh53763 352 FVR--ILRIFKLTR-----HFVG---LRVLGHTLRASTNEFLLLIIFL 389
igfsvigyllfkgyedlse....nev dgnsefssyfdafyflfvltttvG
+ +i++ + ++ e+      +++ +      +++++f +++ +f++++vt+tt+G
Fbh53763 390 ALGVLIFATMIYYAERVGApndPSASEHTQFKNIPIGFWWAVVTMTTLG 439
fGdlvpv.wlgiiffvlfiffiivgl1111111liavi<-*
+Gd++p +w+g++++ ++++++g+l++++++vi
Fbh53763 440 YGDMYPQcWSGMLVG-ALCALAGVLTIAMPVPI 472

```

//

Searching for complete domains in SMART

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /ddm/robison/smart/smart/smart.all.hmms

Sequence file: /prod/ddm/wspace/orfanal/oa-script.4688.seq

Query: Fbh53763

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
BTB_4		72.7	7.8e-18	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
BTB_4	1/1	8	159	1	114	72.7	7.8e-18

Alignments of top-scoring domains:

BTB_4: domain 1 of 1, from 8 to 159: score 72.7, E = 7.8e-18

```

*->cDvtlvvggdlggdnaegkkfhasqHkavLaacrrdSpyFkalfes.
++v+l+vvgg      ++ ++      ++L +      ++ +al+ s+
Fbh53763 8 ERVILNVGG-----TRHET--YRSTLCTL---PGTRLALLASs 40
.....
+++++ ++ +++ ++++++ ++++++ ++ ++++++ ++ ++ +++++
Fbh53763 41 eppgdclttagdklqspplsppprapplspgpggcfeeggagncssrgg 90
.....ieldDeallievspeaFralLnflyt.kldlpeedvenve
+ +++++++D      ++p +F +Ln+++t+kl++p + ++
Fbh53763 91 rasdhpgggrEFFFD-----RHPGVFAYVLNYYRTgKLHCPAD--VCGP 132
elLelAdfldSYGqip.lvelCeefllknk<-*
+ e+ f+ G+ ++ ve C+++ +++
Fbh53763 133 LFEEELAFW---GIDeDVEPCCWMTYRQH 159

```

//

FIGURE 18B

LUSTAL W (1.74) multiple sequence alignment

```

bh53763pat  MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTTAGDKLQSPPPP
atCIKE      MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTAAGDKLQPLPPP
          *****

bh53763pat  LSPPPRAPPLSPGPGGCFEGGAGNCSSRGGRASDHPPGGGREFFFDHRHGPVFAYVLNYYRT
atCIKE      LSPPPRPPLSPVPSGCFEGGAGNCSSHGNGSDHPPGGGREFFFDHRHGPVFAYVLNYYRT
          *****

bh53763pat  GKLHCPADVCGPLFEEELAFWGI DETDVEPCWMTYRQHRDAEEALDIFETPDLIGGDPG
atCIKE      GKLHCPADVCGPLFEEELAFWGI DETDVEPCWMTYRQHRDAEEALDIFETPDLIGGDPG
          *****

bh53763pat  DDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLOPRMWALFEDPYSSRAARFI AFASLFFIL
atCIKE      DDEDLGGKRLGIEDAAGLGGPDGKSGRWKLOPRMWALFEDPYSSRAARFI AFASLFFIL
          *****

bh53763pat  VSITTFCLTM2ETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVTM2CVVWFTFEFLVR
atCIKE      VSITTFCLTM2ETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVTM2CVVWFTFEFLVR
          *****

bh53763pat  IVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAADVLGFLRVVRFVRLRIFK
atCIKE      IVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAADVLGFLRVVRFVRLRIFK
          *****

bh53763pat  LTRHFVGLRVLGHTLRASTNEFLLLIIFLALGVLI FATMIYYAERVGAQPNTM5DPSASEHTQ
atCIKE      LTRHFVGLRVLGHTLRASTNEFLLLIIFLALGVLI FATMIYYAERVGAQPNTM5DPSASEHTQ
          *****

bh53763pat  FKNIPPoreIGFWWAVVTMTTLGYGDMYPQTWSQMLVGALCALAGVLTTM6AMPVPVIVNNFGMY
atCIKE      FKNIPPoreIGFWWAVVTMTTLGYGDMYPQTWSQMLVGALCALAGVLTTM6AMPVPVIVNNFGMY
          *****

bh53763pat  SLAMAKQKLPRKRKKHI PPAPQASSPTFCKTELNMACNSTQSDTCLGKDNRLLEHNRSVL
atCIKE      SLAMAKQKLPRKRKKHI PPAPLASSPTFCKTELNMACNSTQSDTCLGKENRLLEHNRSVL
          *****

bh53763pat  SGDDSTGSEPPLSPPERLPPoreIRRSSTRDKNRRGETCFLTM6LTGDTYTCASDGGIRKGYEKSRS
atCIKE      SGDDSTGSEPPLSPPERLPPoreIRRSSTRDKNRRGETCFLTM6LTGDTYTCASDGGIRKGYEKSRS
          *****

bh53763pat  LNNIAGLAGNALRLSPVTSPYNPCPLRRSRSPIPSIL
atCIKE      LNNIAGLAGNALRLSPVTSPYNPCPLRRSRSPIPSIL
          *****
```

FIGURE 19

Input file Fbh67076FL.seq; Output File Fbh67076FL.tra
Sequence length 6582

```

CCACGCGTCCGCCCACGCGTCCGCCCACGCGTCCGAGAAGGCTTAGGTGGGCAGGCAGGACGAGAGAAAGACTGAGAGG
AGGGAAAGCCGCGTAGGTGGGAGTACAGCGGCGCGAGGGTCGAGGGGGAACCTTCGTCGGTGCAGATGAGGAGGGTGGG
CTTTCAGAACTAGTCCCCCTCGCACCCCGCCCCGCCCCCTCCGCGCTGGGGTCTTCACGGTGCCCTGCCTCAGAGCCG
GGCTCCACCACGCCCCGGAAGAGGGAGTCTGGCCGTCGGCTGGCTCAGGGCGGGCCGGTTGGCTGTACCCAGGCTCCCTG
GCCCCAGTGCGGGACCAGAGCGCGGGGCGGCGCGGCAGCCGCGGGCCGAGGAGGGGCTGCGAGCGAAACGGCGCGGCGC
GGCACGGCGGACGAGTTAGGGCCGGGGCGAGGGAGGCTGTGGCTCCCGACAGAGACAGGGGAGTAGTGTGGGCTGAGG

M F R R S L N 7
CGAGACAGCCCCGGTAGAGCCCAGCTCAGCGCCCCGCGAGCCTTCGACGCG ATG TTC CGC CGG AGC TTG AAT 21

R F C A G E E K R V G T R T V F V G N H 27
CGT TTT TGT GCT GGA GAA GAG AAA CGA GTT GGC ACA CGC ACA GTG TTT GTT GGC AAT CAT 81

P V S E T E A Y I A Q R F C D N R I V S 47
CCA GTT TCG GAA ACA GAA GCT TAC ATT GCA CAA AGA TTT TGT GAT AAT AGA ATA GTC TCA 141

S K Y T L W N F L P K N L F E Q F R R I 67
TCT AAG TAT ACA CTT TGG AAT TTT CTC CCA AAG AAT CTG TTT GAA CAG TTT AGA AGA ATT 201

A N F Y F L I I F L V Q V T V D T P T S 87
GCA AAT TTT TAT TTT CTC ATA ATC TTC CTT GTA CAG GTC ACA GTA GAC ACA CCA ACT AGC 261

P V T S G L P L F F V I T V T A I K Q G 107
CCA GTT ACC AGT GGA CTT CCA CTT TTC TTT GTT ATA ACT GTT ACA GCC ATC AAG CAG GGA 321

Y E D W L R H R A D N E V N K S T V Y I 127
TAT GAG GAT TGG CTG AGA CAC AGA GCT GAC AAT GAA GTC AAC AAA AGC ACT GTT TAC ATT 381

I E N A K R V R K E S E K I K V G D V V 147
ATT GAA AAT GCA AAG CGA GTG AGA AAA GAA AGT GAA AAA ATC AAG GTT GGT GAT GTA GTA 441

E V Q A D E T F P C D L I L L S S C T T 167
GAA GTA CAG GCA GAT GAA ACC TTT CCC TGT GAT CTT ATT CTT CTA TCA TCT TGC ACC ACT 501

D G T C Y V T T A S L D G E S N C K T H 187
GAT GGA ACC TGT TAT GTC ACT ACA GCC AGT CTT GAT GGG GAA TCC AAT TGC AAG ACA CAT 561

Y A V R D T I A L C T A E S I D T L R A 207
TAT GCA GTA CGT GAT ACC ATT GCA CTG TGT ACA GCA GAA TCC ATC GAT ACC CTC CGA GCA 621

A I E C E Q P Q P D L Y K F V G R I N I 227
GCA ATT GAA TGT GAA CAG CCT CAA CCT GAC CTC TAC AAA TTT GTT GGG CGA ATC AAT ATC 681

Y S N S L E A V A R S L G P E N L L L K 247
TAC AGT AAT AGT CTT GAG GCT GTT GCC AGG TCT TTG GGA CCT GAA AAT CTC TTG CTG AAA 741

G A T L K N T E K I Y G V A V Y T G M E 267
GGA GCT ACG CTA AAA AAT ACC GAG AAG ATA TAT GGA GTT GCT GTT TAC ACT GGA ATG GAA 801

```

FIGURE 20A

T	K	M	A	L	N	Y	Q	G	K	S	Q	K	R	S	A	V	E	K	S	287
ACC	AAA	ATG	GCT	TTG	AAC	TAC	CAA	GGG	AAA	TCT	CAG	AAA	CGT	TCT	GCT	GTT	GAA	AAA	TCT	861
I	N	A	F	L	I	V	Y	L	F	I	L	L	T	K	A	A	V	C	T	307
ATT	AAT	GCT	TTC	CTG	ATT	GTA	TAT	TTA	TTT	ATC	TTA	CTG	ACC	AAA	GCT	GCA	GTA	TGC	ACT	921
T	L	K	Y	V	W	Q	S	T	P	Y	N	D	E	P	W	Y	N	Q	K	327
ACT	CTA	AAG	TAT	GTT	TGG	CAA	AGT	ACC	CCA	TAC	AAT	GAT	GAA	CCT	TGG	TAT	AAC	CAA	AAG	981
T	Q	K	E	R	E	T	L	K	V	L	K	M	F	T	D	F	L	S	F	347
ACT	CAG	AAA	GAG	CGA	GAG	ACC	TTG	AAG	GTT	TTA	AAA	ATG	TTC	ACC	GAC	TTC	CTA	TCA	TTT	1041
M	V	L	F	N	F	I	I	P	V	S	M	Y	V	T	V	E	M	Q	K	367
ATG	GTT	CTA	TTC	AAC	TTT	ATC	ATT	CCT	GTC	TCC	ATG	TAC	GTC	ACA	GTA	GAA	ATG	CAG	AAA	1101
F	L	G	S	F	F	I	S	W	D	K	D	F	Y	D	E	E	I	N	E	387
TTC	TTG	GGC	TCC	TTC	TTC	ATC	TCA	TGG	GAT	AAG	GAC	TTT	TAT	GAT	GAA	GAA	ATT	AAT	GAA	1161
G	A	L	V	N	T	S	D	L	N	E	E	L	G	Q	V	D	Y	V	F	407
GGA	GCC	CTG	GTT	AAC	ACA	TCA	GAC	CTT	AAT	GAA	GAA	CTT	GGT	CAG	GTG	GAT	TAT	GTA	TTT	1221
T	D	K	T	G	T	L	T	E	N	S	M	E	F	I	E	C	C	I	D	427
ACA	GAT	AAG	ACT	GGA	ACA	CTC	ACT	GAA	AAC	AGC	ATG	GAA	TTC	ATT	GAA	TGC	TGC	ATA	GAT	1281
G	H	K	Y	K	G	V	T	Q	E	V	D	G	L	S	Q	T	D	G	T	447
GGC	CAC	AAA	TAT	AAA	GGT	GTA	ACT	CAA	GAG	GTT	GAT	GGA	TTA	TCT	CAA	ACT	GAT	GGA	ACT	1341
L	T	Y	F	D	K	V	D	K	N	R	E	E	L	F	L	R	A	L	C	467
TTA	ACA	TAT	TTT	GAC	AAA	GTA	GAT	AAG	AAT	CGA	GAA	GAG	CTG	TTT	CTA	CGT	GCC	TTG	TGT	1401
L	C	H	T	V	E	I	K	T	N	D	A	V	D	G	A	T	E	S	A	487
TTA	TGT	CAT	ACT	GTA	GAA	ATC	AAA	ACA	AAC	GAT	GCT	GTT	GAT	GGA	GCT	ACA	GAA	TCA	GCT	1461
E	L	T	Y	I	S	S	S	P	D	E	I	A	L	V	K	G	A	K	R	507
GAA	TTA	ACC	TAT	ATC	TCC	TCT	TCA	CCA	GAT	GAA	ATA	GCT	TTG	GTG	AAA	GGA	GCT	AAA	AGG	1521
Y	G	F	T	F	L	G	N	R	N	G	Y	M	R	V	E	N	Q	R	K	527
TAC	GGG	TTC	ACA	TTT	TTA	GGA	AAT	CGA	AAT	GGA	TAT	ATG	AGA	GTA	GAG	AAC	CAA	AGA	AAA	1581
E	I	E	E	Y	E	L	L	H	T	L	N	F	D	A	V	R	R	R	M	547
GAA	ATA	GAA	GAA	TAT	GAA	CTT	CTT	CAC	ACC	TTA	AAC	TTT	GAT	GCT	GTC	CGG	CGA	CGT	ATG	1641
S	V	I	V	K	T	Q	E	G	D	I	L	L	F	C	K	G	A	D	S	567
AGT	GTA	ATT	GTG	AAG	ACT	CAA	GAA	GGA	GAC	ATA	CTT	CTC	TTT	TGT	AAA	GGA	GCA	GAC	TCG	1701
A	V	F	P	R	V	Q	N	H	E	I	E	L	T	K	V	H	V	E	R	587
GCA	GTT	TTT	CCC	AGA	GTG	CAA	AAT	CAT	GAA	ATT	GAG	TTA	ACT	AAA	GTC	CAT	GTG	GAA	CGT	1761
N	A	M	D	G	Y	R	T	L	C	V	A	F	K	E	I	A	P	D	D	607
AAT	GCA	ATG	GAT	GGG	TAT	CGG	ACA	CTC	TGT	GTA	GCC	TTC	AAA	GAA	ATT	GCT	CCA	GAT	GAT	1821
Y	E	R	I	N	R	Q	L	I	E	A	K	M	A	L	Q	D	R	E	E	627
TAT	GAA	AGA	ATT	AAC	AGA	CAG	CTC	ATA	GAG	GCA	AAA	ATG	GCC	TTA	CAA	GAC	AGA	GAA	GAA	1881
K	M	E	K	V	F	D	D	I	E	T	N	M	N	L	I	G	A	T	A	647
AAA	ATG	GAA	AAA	GTT	TTC	GAT	GAT	ATT	GAG	ACA	AAC	ATG	AAT	TTA	ATT	GGA	GCC	ACT	GCA	1941

FIGURE 20B

V	E	D	K	L	Q	D	Q	A	A	E	T	I	E	A	L	H	A	A	G	667
GTT	GAA	GAC	AAG	CTA	CAA	GAT	CAA	GCT	GCA	GAG	ACC	ATT	GAA	GCT	CTG	CAT	GCA	GCA	GGC	2001
L	K	V	W	V	L	T	G	D	K	M	E	T	A	K	S	T	C	Y	A	687
CTG	AAA	GTC	TGG	GTG	CTC	ACT	GGG	GAC	AAG	ATG	GAG	ACA	GCT	AAA	TCC	ACA	TGC	TAT	GCC	2061
C	R	L	F	Q	T	N	T	E	L	L	E	L	T	T	K	T	I	E	E	707
TGC	CGC	CTT	TTC	CAG	ACC	AAC	ACT	GAG	CTC	TTA	GAA	CTA	ACC	ACA	AAA	ACC	ATT	GAA	GAA	2121
S	E	R	K	E	D	R	L	H	E	L	L	I	E	Y	R	K	K	L	L	727
AGT	GAA	AGG	AAA	GAA	GAT	CGA	TTA	CAT	GAA	TTA	TTG	ATA	GAA	TAT	CGC	AAG	AAA	TTG	CTG	2181
H	E	F	P	K	S	T	R	S	F	K	K	A	W	T	E	H	Q	E	Y	747
CAT	GAG	TTT	CCT	AAA	AGT	ACT	AGA	AGC	TTT	AAA	AAA	GCA	TGG	ACA	GAA	CAT	CAG	GAA	TAT	2241
G	L	I	I	D	G	S	T	L	S	L	I	L	N	S	S	Q	D	S	S	767
GGA	TTA	ATC	ATA	GAT	GGC	TCC	ACA	TTG	TCA	CTC	ATA	CTA	AAT	TCT	AGT	CAA	GAC	TCT	AGT	2301
S	N	N	Y	K	S	I	F	L	Q	I	C	M	K	C	T	A	V	L	C	787
TCA	AAC	AAT	TAC	AAA	AGC	ATT	TTC	CTA	CAA	ATA	TGT	ATG	AAG	TGT	ACT	GCA	GTG	CTC	TGC	2361
C	R	M	A	P	L	Q	K	A	Q	I	V	R	M	V	K	N	L	K	G	807
TGT	CGG	ATG	GCA	CQA	TTA	CAG	AAA	GCC	CAG	ATT	GTC	AGA	ATG	GTG	AAG	AAT	TTA	AAA	GGC	2421
S	P	I	T	L	S	I	G	D	G	A	N	D	V	S	M	I	L	E	S	827
AGC	CCA	ATA	ACT	CTG	TCG	ATA	GGT	GAT	GGT	GCC	AAT	GAT	GTT	AGT	ATG	ATC	TTG	GAA	TCC	2481
H	V	G	I	G	I	K	G	K	E	G	R	Q	A	A	R	N	S	D	Y	847
CAT	GTG	GGA	ATA	GGT	ATT	AAA	GGC	AAA	GAA	GGT	CGC	CAA	GCA	GCT	AGG	AAT	AGC	GAT	TAT	2541
S	V	P	K	F	K	H	L	K	K	L	L	L	A	H	G	H	L	Y	Y	867
TCT	GTT	CCA	AAG	TTT	AAA	CAC	TTA	AAG	AAA	CTG	CTG	TTG	GCT	CAT	GGA	CAT	CTA	TAT	TAT	2601
V	R	I	A	H	L	V	Q	Y	F	F	Y	K	N	L	C	F	I	L	P	887
GTG	AGA	ATA	GCA	CAC	CTT	GTA	CAG	TAC	TTC	TTC	TAT	AAG	AAC	CTT	TGT	TTC	ATT	TTG	CCA	2661
Q	F	L	Y	Q	F	F	C	G	F	S	Q	Q	P	L	Y	D	A	A	Y	907
CAG	TTT	TTG	TAC	CAG	TTC	TTC	TGT	GGA	TTC	TCA	CAA	CAG	CCA	CTG	TAT	GAT	GCT	GCT	TAC	2721
L	T	M	Y	N	I	C	F	T	S	L	P	I	L	A	Y	S	L	L	E	927
CTT	ACA	ATG	TAC	AAT	ATC	TGC	TTC	ACA	TCC	TTG	CCC	ATC	CTG	GCC	TAT	AGT	CTA	CTG	GAA	2781
Q	H	I	N	I	D	T	L	T	S	D	P	R	L	Y	M	K	I	S	G	947
CAG	CAC	ATC	AAC	ATT	GAC	ACT	CTG	ACC	TCA	GAT	CCC	CGA	TTG	TAT	ATG	AAA	ATT	TCT	GGC	2841
N	A	M	L	Q	L	G	P	F	L	Y	W	T	F	L	A	A	F	E	G	967
AAT	GCC	ATG	CTA	CAG	TTG	GGC	CCC	TTC	TTA	TAT	TGG	ACA	TTT	CTG	GCT	GCC	TTT	GAA	GGG	2901
T	V	F	F	F	G	T	Y	F	L	F	Q	T	A	S	L	E	E	N	G	987
ACA	GTG	TTC	TTC	TTT	GGG	ACT	TAC	TTT	CTT	TTT	CAG	ACT	GCA	TCC	CTA	GAA	GAA	AAT	GGA	2961
K	V	Y	G	N	W	T	F	G	T	I	V	F	T	V	L	V	F	T	V	1007
AAG	GTA	TAC	GGA	AAC	TGG	ACT	TTT	GGA	ACC	ATT	GTT	TTT	ACA	GTC	TTA	GTA	TTC	ACT	GTA	3021
T	L	K	L	A	L	D	T	R	F	W	T	W	I	N	H	F	V	I	W	1027
ACC	CTG	AAG	CTT	GCC	TTG	GAT	ACC	CGA	TTC	TGG	ACG	TGG	ATA	AAT	CAC	TTT	GTG	ATT	TGG	3081

FIGURE 20C

G S L A F Y V F F S F F W G G I I W P F 1047
GGT TCT TTA GCC TTC TAT GTA TTT TTC TCA TTC TTC TGG GGA GGA ATT ATT TGG CCT TTT 3141
L K Q Q R M Y F V F A Q M L S S V S T W 1067
CTC AAG CAA CAG AGA ATG TAT TTT GTA TTT GCC CAA ATG CTG TCT TCT GTA TCC ACA TGG 3201
L A I I L L I F I S L F P E I L L I V L 1087
TTG GCT ATA ATT CTT CTA ATA TTT ATC AGC CTG TTC CCT GAG ATT CTT CTG ATA GTA TTA 3261
K N V R R R S A R R N L S C R R A S D S 1107
AAG AAT GTA AGA AGA AGA AGT GCC AGG AGA AAT CTG AGC TGT AGA AGG GCA TCT GAC TCA 3321
L S A R P S V R P L L L R T F S D E S N 1127
TTA TCC GCC AGA CCT TCA GTC AGA CCT CTT CTT TTA CGA ACA TTC TCA GAC GAA TCT AAT 3381
V L 1130
GTA TTG TAA 3390

CAGAATCCGAATCTTGAAGTGCCTATGTTATTGTCTACAAGCATACTGACAGTGGTTACAGCTAAAAAGAAAGCATG
AAGAAACAACACTACAAAAAGTTATCATCTCAGGATACTTGATATGCAACACACTAAACCACTCTCATGTCTAGAGTTCAC
AATAAATGTTTCATTAAAAATACCAAATGATTCTCTTAAGCATTTACCATTATTGTAAGTAGCCTTTATGGCCAAAGCTGT
AAGTTAAGAATTATATGAAAGTTGAAAGCAAGAATACTTAGAATTCTGGCTTTAGTTAGAGTAATATAACTCAAATGGG
TGCTCTTTTAACCCATGAACCTTTGTGAATGGATTTAAATACAATAGTATGAAGTAGAAGTTATGCAATGAGAATGAATA
GATTTTGCTAATACTACTTTTTTTGCCTGGCAGAAGAAATAGACTATTTGGATCACATTTCTCATTCTCCTAAATGAT
CATCTTAATTTTTTTTCCCAAGTACATAAGGAATACTTGAAAATACAGAATAACTAAATAGTATCAATGCATCAGACAG
AATAGTTAATCCCTTCTGTTTACCCATGTGCTACTAATGTCTTGGTAGAATATTCTTGCCAAAAAATACCTTGAACGC
TTATGTGGAAAGTGTTAACTTACGGGTATTTTTGTGGGAATAGAAAAAATGTTTATTTTTTTATTCTTCTGAATTAAA
CCCCACTTATGGGTGTAAGCCTACTAGACTTGAAAATAAAGTATAAAACATTTCCAATCACTTAGTAGCCCCCTCAAAGT
AGTTAGAAAAATAAACAGATTTTCCAGTGTGATTTTACTGGGATCTGCAGTAAGGTGGTTTAAACCATAGTTATATAA
AAATAAAGGTCACTTCTGAATATCAGCCTTTTATAATTTTATGTGAAGAGGAAGAAATATAGCTTATTTTAACTTTTGA
CGGTTTTTATTGTAAAGAGATTGCATTTATGCATATATGCAGTGCTTTTTCTTAAACTTGCCCAATTTGGAAAGGGGA
AGGAGCCACCCCAAACGGTGGTTCAGCTTGTAGAGCCATGACTCTGTGAAGATGAATGTTGTCTCTTAACCTGGACAG
GGAAATGGTCTAACTCTAAACCATGTAAGTACCTTAGTAAAGTCCTTGACTAACTGAAGTGAAGGAAGGTTTAGCCT
TCTAATTAGTTCACCTTGAAACATAAATGTGAATGTCTTCAATCAATGTTAAACACATACTTTTTTGGATATAAATGAC
CATATTTATTTGACTGCTAGTTTTTTTTGTCTTTCTGGCATGCCTGTACTATTATTAATGTTTATATTG
TACCTTGATTTGGAAAAGTATTGGAGTTAATCTGTATTATATTTATATAGTCCATATGGCACATTTGATTCTTCCACAT
ATATTTTGTGTTAATGTTTAGGTATGATTTTTTTCTAAATCTAGAAAAGAACATAATTCAGTTATCAGAAGCCATTC
CATCATTATAGACCCTTTTTCATTATTTTCAATTTGCTCTCATATATCAGTATTATTTTTGAGCATTTTGTTACATGTCAT

FIGURE 20D

TCACAACTTACCTAAGTGTGCTGTGTTCTGGTAGCCCGTATTTGAGGTAAGCTGCTGAAAACAAAAGTCTCTATATTCT
TTGCCTATTCCAAAGAGCTAAAAAGTCTAACCAGGAAAGCTTTTGATATTTTGTGTTTGTCTTGTCTTATGGT
TGTGTGTGCTGTATTATGATTGCTGTTTTACATAAAATCTATGGGAAGTGTGAATACAGACAAGAGAGCCACAGTAGAG
AGGCTTGTMTAATGCAGTACCATGGAGAGTTAACAGAATAATCTAGTAGAAAAATAACTGGTTGCATGTAAAATTCCT
TCCAGCCAGAAAGAAAGAAAGACAAGGAGTAAGGGGGATTTAGAGTTATGTCTCAGCTACACATTACATTGTGATACTG
CAGCTCAAATTCAGAATGGCAATGATACATGATATCATGGCCTAGATCCTTGAGAGGGACCTGGCTTTCCTTTTTTAAAA
GATATTTTACTGAAGAGCTAAAACTGGCCAGTGTGGGGTTAGCAGATCGAATAACTTGAAATAGACCGTGCAGTATTC
CTAGCACTCAATGTAATCACCCCTATTTGTGACAGAGAAAGGGAAAAAATATAATAAGATCATCTACCTATAATTTGAA
TAATTTTGAGCTATCAAAATGTCTTTGTAATTTTCACAACCGCTGTCCATTGTTTGAGGATGTTACCTACTAACTGAA
AACATTTCATTCCATATCTACTTACACATACACCAGCAACAGTATAAATGTAAGCCTAACTTTGCAAAATTCGTAATAAT
TTAGTGATGGAATTTTTTAATAACATGCAGTATATAAATGTGCAGATTTTATGCGTGTTGACAAAATCATTTTTTCAGCT
TGCAAAATGGGACTGCAATATTACATTTTTCACTTAAGCAGTTTTTTACATCTACGTTGTTGCTTCTAAAATGAATGT
GAATGCCATCTTTTATGACTGCAACTTGCCTTTTCCATTACAGAAATTTTTGTTTGATGTAATCAATAAACTTTGGTAT
GATAAA

FIGURE 20E

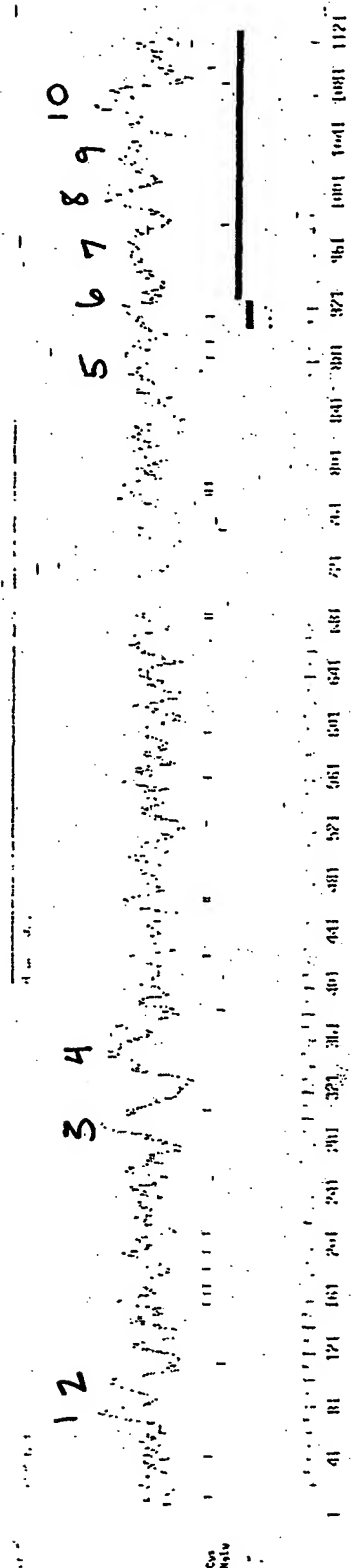


FIGURE 21

Protein Family / Domain Matches, HMMer version 2
 Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)
 Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.13758.seq

Query: 67076

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Hydrolase 1	haloacid dehalogenase-like hydrolase	12.7	0.019	

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
Hydrolase	1/1	403	837	1	184 [1]	12.7	0.019

Alignments of top-scoring domains:

Hydrolase: domain 1 of 1, from 403 to 837: score 12.7, E = 0.019

```

*->ikavvFDkDGTLTdgkeppiaaeivealrelgl.....
++ v+ Dk+GTLt+ + e +++ +g+++ ++ ++++++
67076 403 VDYVFTDKTGTLTEN-SMEFIECCIDGHKYKGVtqevdglstqtdgtl 448

....apleeveklgrglgerilleggltaell.....
+ + + + e l +r+l + + + + t + + + + + + + + + +
67076 449 tyfdKVDKNREELFLRALcL--CHTVEIKTNDVdgatesaeltysssp 496

.....
++ ++ + + + + + + + + + + + + + + + + + +
67076 497 deialvkgakrygftflgnrngymrvenqrkeieeyellhtlnfdavrrr 546

.....
+ ++ + + + + + + + + + + + + + + + + + + +
67076 547 msvivktqegdillfckgadsavfprvqnheiltkvhvernarnamdyrtl 596

.....ld.evlgli
++ + + + + + + + + + + + + + + + + + + + +
67076 597 cvafkeiapddyerinrqlieakmalqdreemkvfddieTNmNLIGAT 646

al.dklypgarealkaLkerGikvailTngdr.naealle.....algla
a++dkl + a++++aL+++G+kv++lT++ + + + + + + + +
67076 647 AVeDKLQDQAETIEALHAAGLKVWVLTGDKMeTAKSTCYacrlfQTNTE 696

.lfdaivdsdevggvpgvppvvgKPkpeifllalerlgvkpeevg.....
l + + + + + + + + + + + + + + + + + + + +
67076 697 lLELTTKTIEESE-----RKEDRLHELLIEYRKLL--Hefpkstr 735

.....
+ ++ + + + + + + + + + + + + + + + + + + +
67076 736 sfkkawtehqeygliidgstlsilnssqdssnnyksiflqicmkctav 785

.....p.kvlmvGDginDapalaaAGvgvamgn
+ + + + + + + + + + + + + + + + + + + +
67076 786 lccrmaplqkaqivrmvknlgSpITLSIGDGANDVSMILESHVGIGIKG 835

gg<-*
67076 836 KE 837

```

FIGURE 22

ILAHVGIGVGIGKEGRQAARNSDYAIPKFKHLKKMLLVHGHFYIRISELVQYFFYKNVC

FILPQFLYQHF C G F S Q Q P L Y D A A Y L T M Y N I C F T S L P I L A Y S L L E Q H I N I D T L T S D P R L Y M
F I F P Q F L Y Q H F C G F S Q Q T L Y D T A Y L T L Y N I S F T S L P I L L Y S L M E Q H V G I D V L K R D P T L Y R

KISGNAMLQLGFLYWTFLAAFEGETVFFPGTYFLEQTASLEENGKVYGNWTFGTIVFTVL
DIAKNALLRWRFIYWTFGLGVFDALVEFFGAYFI FENTTVTINGOMFGNWTEGTLVETVM

* * * * *

VFTVTU[KLALDTRFWTWI]NHFVIWGS LAPYVFFS FFWG GI IWPFLKQORMYFVFAQMLSS
VLTVTLKLALDTHYWTWI]NHFVIWGS LLFYIAFSLLWG GVIWPFLSYQRMYYVFIS MLSS

.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.

VSTWLAIIILLIFISLFPÉILLIVIK-----NVRRRSARRNLSRRASDLSAR
GPAWLGIILLVTVGLLPDVLKVKVLCROLWPTATERTONIQHODSISEFTPLASLPWGAQ
* *

-----PSVRPLLLRTFSDES NVL-----
 GSRLLAACSSPSGRVVC SRWESEECFVLP LHPGLPHKARYGCCRSSLEMP
 * * * : * * * *

41/79

Input file Fbh67102FL.seq; Output File Fbh67102FL.tra
Sequence length 6074

```

CCACGCGTCCGGGAGGAGCGGAGGGAGAAGTAGGTTGCGAGCTCAGCACAGGCTCCGGCGCTGGCTCCCGCAGCTGAGT
TTGGGAGATGTCTAAGTGATTTTTTTTTTTTCCCGGAAGGCAATGGCTGGCGTGGGAAGCACAAACCCGCTTTTCACTCTT
CGAATTTGTGCTTAGCTCTTTTCTTGACCTTGGGACTCGTGACCAACATGCTGTGATGTGTGCCGAGGGAGGAATTGG

                                M T E A L Q W A R Y
TCAGCTACACAACCTGGATCTTACCACAGTTTGAT ATG ACT GAG GCT CTC CAA TGG GCC AGA TAT 10
                                30
H W R R L I R G A T R D D D S G P Y N Y 30
CAC TGG CGA CGG CTG ATC AGA GGT GCA ACC AGG GAT GAT GAT TCA GGG CCA TAC AAC TAT 90
S S L L A C G R K S S Q I P K L S G R H 50
TCC TCG TTG CTC GCC TGT GGG CGC AAG TCC TCT CAG ATC CCT AAA CTG TCA GGA AGG CAC 150
R I V V P H I Q P F K D E Y E K F S G A 70
CGG ATT GTT GTT CCC CAC ATC CAG CCC TTC AAG GAT GAG TAT GAG AAG TTC TCC GGA GCC 210
Y V N N R I R T T K Y T L L N F V P R N 90
TAT GTG AAC AAT CGA ATA CGA ACA ACA AAG TAC ACA CTT CTG AAT TTT GTG CCA AGA AAT 270
L F E Q F H R A A S L Y F L F L V V L N 110
TTA TTT GAA CAA TTT CAC AGA GCT GCC AGT TTA TAT TTC CTG TTC CTA GTT GTC CTG AAC 330
W V P L V E A F Q K E I T M L P L V V V 130
TGG GTA CCT TTG GTA GAA GCC TTC CAA AAG GAA ATC ACC ATG TTG CCT CTG GTG GTG GTC 390
L T I I A I K D G L E D Y R K Y K I D K 150
CTT ACA ATT ATC GCA ATT AAA GAT GGC CTG GAA GAT TAT CGG AAA TAC AAA ATT GAC AAA 450
Q I N N L I T K V Y S R K E K K Y I D R 170
CAG ATC AAT AAT TTA ATA ACT AAA GTT TAT AGT AGG AAA GAG AAA AAA TAC ATT GAC CGA 510
C W K D V T V G D F I R L S C N E V I P 190
TGC TGG AAA GAC GTT ACT GTT GGG GAC TTT ATT CGC CTC TCC TGC AAT GAG GTC ATC CCT 570
A D M V L L F S T D P D G I C H I E T S 210
GCA GAC ATG GTA CTA CTC TTT TCC ACT GAT CCA GAT GGA ATC TGT CAC ATT GAG ACT TCT 630
G L D G E S N L K Q R Q V V R G Y A E Q 230
GGT CTT GAT GGA GAG AGC AAT TTA AAA CAG AGG CAG GTG GTT CGG GGA TAT GCA GAA CAG 690
D S E V D P E K F S S R I E C E S P N N 250
GAC TCT GAA GTT GAT CCT GAG AAG TTT TCC AGT AGG ATA GAA TGT GAA AGC CCA AAC AAT 750
D L S R F R G F L E H S N K E R V G L S 270
GAC CTC AGC AGA TTC CGA GGC TTC CTA GAA CAT TCC AAC AAA GAA CGC GTG GGT CTC AGT 810
K E N L L L R G C T I R N T E A V V G I 290
AAA GAA AAT TTG TTG CTT AGA GGA TGC ACC ATT AGA AAC ACA GAG GCT GTT GTG GGC ATT 870
V V Y A G H E T K A M L N N S G P R Y K 310
GTG GTT TAT GCA GGC CAT GAA ACC AAA GCA ATG CTG AAC AAC AGT GGG CCA CGG TAT AAG 930

```

FIGURE 24A

R	S	K	L	E	R	R	A	N	T	D	V	L	W	C	V	M	L	L	V	330
CGC	AGC	AAA	TTA	GAA	AGA	AGA	GCA	AAC	ACA	GAT	GTC	CTC	TGG	TGT	GTC	ATG	CTT	CTG	GTC	990
I	M	C	L	T	G	A	V	G	H	G	I	W	L	S	R	Y	E	K	M	350
ATA	ATG	TGC	TTA	ACT	GGC	GCA	GTA	GGT	CAT	GGA	ATC	TGG	CTG	AGC	AGG	TAT	GAA	AAG	ATG	1050
H	F	F	N	V	P	E	P	D	G	H	I	I	S	P	L	L	A	G	F	370
CAT	TTT	TTC	AAT	GTT	CCC	GAG	CCT	GAT	GGA	CAT	ATC	ATA	TCA	CCA	CTG	TTG	GCA	GGA	TTT	1110
Y	M	F	W	T	M	I	I	L	L	Q	V	L	I	P	I	S	L	Y	V	390
TAT	ATG	TTT	TGG	ACC	ATG	ATC	ATT	TTG	TTA	CAG	GTC	TTG	ATT	CCT	ATT	TCT	CTC	TAT	GTT	1170
S	I	E	I	V	K	L	G	Q	I	Y	F	I	Q	S	D	V	D	F	Y	410
TCC	ATC	GAA	ATT	GTG	AAG	CTT	GGA	CAA	ATA	TAT	TTC	ATT	CAA	AGT	GAT	GTG	GAT	TTC	TAC	1230
N	E	K	M	D	S	I	V	Q	C	R	A	L	N	I	A	E	D	L	G	430
AAT	GAA	AAA	ATG	GAT	TCT	ATT	GTT	CAG	TGC	CGA	GCC	CTG	AAC	ATC	GCC	GAG	GAT	CTG	GGA	1290
Q	I	Q	Y	L	F	S	D	K	T	G	T	L	T	E	N	K	M	V	F	450
CAG	ATT	CAG	TAC	CTC	TTT	TCC	GAT	AAG	ACA	GGA	ACC	CTC	ACT	GAG	AAT	AAG	ATG	GTT	TTT	1350
R	R	C	S	V	A	G	F	D	Y	C	H	E	E	N	A	R	R	L	E	470
CGA	AGA	TGT	AGT	GTG	GCA	GGA	TTT	GAT	TAC	TGC	CAT	GAA	GAA	AAT	GCC	AGG	AGG	TTG	GAG	1410
S	Y	Q	E	A	V	S	E	D	E	D	F	I	D	T	V	S	G	S	L	490
TCC	TAT	CAG	GAA	GCT	GTC	TCT	GAA	GAT	GAA	GAT	TTT	ATA	GAC	ACA	GTC	AGT	GGT	TCC	CTC	1470
S	N	M	A	K	P	R	A	P	S	C	R	T	V	H	N	G	P	L	G	510
AGC	AAT	ATG	GCA	AAA	CCG	AGA	GCC	CCC	AGC	TGC	AGG	ACA	GTT	CAT	AAT	GGG	CCT	TTG	GGA	1530
N	K	P	S	N	H	L	A	G	S	S	F	T	L	G	S	G	E	G	A	530
AAT	AAG	CCC	TCA	AAT	CAT	CTT	GCT	GGG	AGC	TCT	TTT	ACT	CTA	GGA	AGT	GGA	GAA	GGA	GCC	1590
S	E	V	P	H	S	R	Q	A	A	F	S	S	P	I	E	T	D	V	V	550
AGT	GAA	GTG	CCT	CAT	TCC	AGA	CAG	GCT	GCT	TTC	AGT	AGC	CCC	ATT	GAA	ACA	GAC	GTG	GTA	1650
P	D	T	R	L	L	D	K	F	S	Q	I	T	P	R	L	F	M	P	L	570
CCA	GAC	ACC	AGG	CTT	TTA	GAC	AAA	TTT	AGT	CAG	ATT	ACA	CCT	CGG	CTC	TTT	ATG	CCA	CTA	1710
D	E	T	I	Q	N	P	P	M	E	T	L	Y	I	I	D	F	F	I	A	590
GAT	GAG	ACC	ATC	CAA	AAT	CCA	CCA	ATG	GAA	ACT	TTG	TAC	ATT	ATC	GAC	TTT	TTC	ATT	GCA	1770
L	A	I	C	N	T	V	V	V	S	A	P	N	Q	P	R	Q	K	I	R	610
TTG	GCA	ATT	TGC	AAC	ACA	GTA	GTG	GTT	TCT	GCT	CCT	AAC	CAA	CCC	CGA	CAA	AAG	ATC	AGA	1830
H	P	S	L	G	G	L	P	I	K	S	L	E	E	I	K	S	L	F	Q	630
CAC	CCT	TCA	CTG	GGG	GGG	TTG	CCC	ATT	AAG	TCT	TTG	GAA	GAG	ATT	AAA	AGT	CTT	TTC	CAG	1890
R	W	S	V	R	R	S	S	S	P	S	L	N	S	G	K	E	P	S	S	650
AGA	TGG	TCT	GTC	CGA	AGA	TCA	AGT	TCT	CCA	TCG	CTT	AAC	AGT	GGG	AAA	GAG	CCA	TCT	TCT	1950
G	V	P	N	A	F	V	S	R	L	P	L	F	S	R	M	K	P	A	S	670
GGA	GTT	CCA	AAC	GCC	TTT	GTG	AGC	AGA	CTC	CCT	CTC	TTT	AGT	CGA	ATG	AAA	CCA	GCT	TCA	2010
P	V	E	E	E	V	S	Q	V	C	E	S	P	Q	C	S	S	S	S	A	690
CCT	GTG	GAG	GAA	GAG	GTC	TCC	CAG	GTG	TGT	GAG	AGC	CCC	CAG	TGC	TCC	AGT	AGT	TCA	GCT	2070

FIGURE 24B

C	C	T	E	T	E	K	Q	H	G	D	A	G	L	L	N	G	K	A	E	710
TGC	TGC	ACA	GAA	ACA	GAG	AAA	CAA	CAC	GGT	GAT	GCA	GGC	CTC	CTG	AAT	GGC	AAG	GCA	GAG	2130
S	L	P	G	Q	P	L	A	C	N	L	C	Y	E	A	E	S	P	D	E	730
TCC	CTC	CCT	GGA	CAG	CCA	TTG	GCC	TGC	AAC	CTG	TGT	TAT	GAG	GCC	GAG	AGC	CCA	GAC	GAA	2190
A	A	L	V	Y	A	A	R	A	Y	Q	C	T	L	R	S	R	T	P	E	750
GCG	GCC	TTA	GTG	TAT	GCC	GCC	AGG	GCT	TAC	CAA	TGC	ACT	TTA	CGG	TCT	CGG	ACA	CCA	GAG	2250
Q	V	M	V	D	F	A	A	L	G	P	L	T	F	Q	L	L	H	I	L	770
CAG	GTC	ATG	GTG	GAC	TTT	GCT	GCT	TTG	GGA	CCA	TTA	ACA	TTT	CAA	CTC	CTA	CAC	ATC	CTG	2310
P	F	D	S	V	R	K	R	M	S	V	V	V	R	H	P	L	S	N	Q	790
CCC	TTT	GAC	TCA	GTA	AGA	AAA	AGA	ATG	TCT	GTT	GTG	GTC	CGA	CAC	CCT	CTT	TCC	AAT	CAA	2370
V	V	V	Y	T	K	G	A	D	S	V	I	M	E	L	L	S	V	A	S	810
GTT	GTG	GTG	TAT	ACG	AAA	GGC	GCT	GAT	TCT	GTG	ATC	ATG	GAG	TTA	GTG	TCG	GTG	GCT	TCC	2430
P	D	G	A	S	L	E	K	Q	Q	M	I	V	R	E	K	T	Q	K	H	830
CCA	GAT	GGA	GCA	AGT	CTG	GAG	AAA	CAA	CAG	ATG	ATA	GTA	AGG	GAG	AAA	ACC	CAG	AAG	CAC	2490
L	D	D	Y	A	K	Q	G	L	R	T	L	C	I	A	K	K	V	M	S	850
TTG	GAT	GAC	TAT	GCC	AAA	CAA	GGC	CTT	CGT	ACT	TTA	TGT	ATA	GCA	AAG	AAG	GTC	ATG	AGT	2550
D	T	E	Y	A	E	W	L	R	N	H	F	L	A	E	T	S	I	D	N	870
GAC	ACT	GAA	TAT	GCA	GAG	TGG	CTG	AGG	AAT	CAT	TTT	TTA	GCT	GAA	ACC	AGC	ATT	GAC	AAC	2610
R	E	E	L	L	L	E	S	A	M	R	L	E	N	K	L	T	L	L	G	890
AGG	GAA	GAA	TTA	CTA	CTT	GAA	TCT	GCC	ATG	AGG	TTG	GAG	AAC	AAA	CTT	ACA	TTA	CTT	GGT	2670
A	T	G	I	E	D	R	L	Q	E	G	V	P	E	S	I	E	A	L	H	910
GCT	ACT	GGC	ATT	GAA	GAC	CGT	CTG	CAG	GAG	GGA	GTC	CCT	GAA	TCT	ATA	GAA	GCT	CTT	CAC	2730
K	A	G	I	K	I	W	M	L	T	G	D	K	Q	E	T	A	V	N	I	930
AAA	GCG	GGC	ATC	AAG	ATC	TGG	ATG	CTG	ACA	GGG	GAC	AAG	CAG	GAG	ACA	GCT	GTC	AAC	ATA	2790
A	Y	A	C	K	L	L	E	P	D	D	K	L	F	I	L	N	T	Q	S	950
GCT	TAT	GCA	TGC	AAA	CTA	CTG	GAG	CCA	GAT	GAC	AAG	CTT	TTT	ATC	CTC	AAT	ACC	CAA	AGT	2850
K	D	A	C	G	M	L	M	S	T	I	L	K	E	L	Q	K	K	T	Q	970
AAA	GAT	GCC	TGT	GGG	ATG	CTG	ATG	AGC	ACA	ATT	TTG	AAA	GAA	CTT	CAG	AAG	AAA	ACT	CAA	2910
A	L	P	E	Q	V	S	L	S	E	D	L	L	Q	P	P	V	P	R	D	990
GCC	CTG	CCA	GAG	CAA	GTG	TCA	TTA	AGT	GAA	GAT	TTA	CTT	CAG	CCT	CCT	GTC	CCC	CGG	GAC	2970
S	G	L	R	A	G	L	I	I	T	G	K	T	L	E	F	A	L	Q	E	1010
TCA	GGG	TTA	CGA	GCT	GGA	CTC	ATT	ATC	ACT	GGG	AAG	ACC	CTG	GAG	TTT	GCC	CTG	CAA	GAA	3030
S	L	Q	K	Q	F	L	E	L	T	S	W	C	Q	A	V	V	C	C	R	1030
AGT	CTG	CAA	AAG	CAG	TTC	CTG	GAA	CTG	ACA	TCT	TGG	TGT	CAA	GCT	GTG	GTC	TGC	TGC	CGA	3090
A	T	P	L	Q	K	S	E	V	V	K	L	V	R	S	H	L	Q	V	M	1050
GCC	ACA	CCG	CTG	CAG	AAA	AGT	GAA	GTG	GTG	AAA	TTG	GTC	CGC	AGC	CAT	CTC	CAG	GTG	ATG	3150
T	L	A	I	G	D	G	A	N	D	V	S	M	I	Q	V	A	D	I	G	1070
ACC	CTT	GCT	ATT	GGT	GAT	GGT	GCC	AAT	GAT	GTT	AGC	ATG	ATA	CAA	GTG	GCA	GAC	ATT	GGG	3210

FIGURE 24C

I G V S G Q E G M Q A V M A S D F A V S 1090
ATA GGG GTC TCA GGT CAA GAA GGC ATG CAG GCT GTG ATG GCC AGT GAC TTT GCC GTT TCT 3270

Q F K H L S K L L L V H G H W C Y T R L 1110
CAG TTC AAA CAT CTC AGC AAG CTC CTT CTT GTC CAT GGA CAC TGG TGT TAT ACA CGG CTT 3330

S N M I L Y F F Y K N V A Y V N L L F W 1130
TCC AAC ATG ATT CTC TAT TTT TTC TAT AAG AAT GTG GCG TAT GTG AAC CTC CTT TTC TGG 3390

Y Q F F C G F S G T S M T D Y W V L I F 1150
TAC CAG TTC TTT TGT GGA TTT TCA GGA ACA TCC ATG ACT GAT TAC TGG GTT TTG ATC TTC 3450

F N L L F T S A P P V I Y G V L E K D V 1170
TTC AAC CTC CTC TTC ACA TCT GCC CCT CCT GTC ATT TAT GGT GTT TTG GAG AAA GAT GTG 3510

S A E T L M Q L P E L Y R S G Q K S E A 1190
TCT GCA GAG ACC CTC ATG CAA CTG CCT GAA CTT TAC AGA AGT GGT CAG AAA TCA GAG GCA 3570

Y L P H T F W I T L L D A F Y Q S L V C 1210
TAC TTA CCC CAT ACC TTC TGG ATC ACC TTA TTG GAT GCT TTT TAT CAA AGC CTG GTC TGC 3630

F F V P Y F T Y Q G S D T D I F A F G N 1230
TTC TTT GTG CCT TAT TTT ACC TAC CAG GGC TCA GAT ACT GAC ATC TTT GCA TTT GGA AAC 3690

P L N T A A L F I V L L H L V I E S K S 1250
CCC CTG AAC ACA GCC GCT CTG TTC ATC GTT CTC CTC CAT CTG GTC ATT GAA AGC AAG AGT 3750

L T W I H L L V I I G S I L S Y F L F A 1270
TTG ACT TGG ATT CAC TTG CTG GTC ATC ATT GGT AGC ATC TTG TCT TAT TTT TTA TTT GCC 3810

I V F G A M C V T C N P P S N P Y W I M 1290
ATA GTT TTT GGA GCC ATG TGT GTA ACT TGC AAC CCA CCA TCC AAC CCT TAC TGG ATT ATG 3870

Q E H M L D P V F Y L V C I L T T S I A 1310
CAG GAG CAC ATG CTG GAT CCA GTA TTC TAC TTA GTT TGT ATC CTC ACG ACG TCC ATT GCT 3930

L L P R F V Y R V L Q G S L F P S P I L 1330
CTT CTG CCC AGG TTT GTA TAC AGA GTT CTT CAG GGA TCC CTG TTT CCA TCT CCA ATT CTG 3990

R A K H F D R L T P E E R T K A L K K W 1350
AGA GCT AAG CAC TTT GAC AGA CTA ACT CCA GAG GAG AGG ACT AAA GCT CTC AAG AAG TGG 4050

R G A G K M N Q V T S K Y A N Q S A G K 1370
AGA GGG GCT GGA AAG ATG AAT CAA GTG ACA TCA AAG TAT GCT AAC CAA TCA GCT GGC AAG 4110

S G R R P M P G P S A V F A M K S A T S 1390
TCA GGA AGA AGA CCC ATG CCT GGC CCT TCT GCT GTA TTT GCA ATG AAG TCA GCA ACT TCC 4170

C A I E Q G N L S L C E T A L D Q G Y S 1410
TGT GCT ATT GAG CAA GGA AAC TTA TCT CTG TGT GAA ACT GCT TTA GAT CAA GGC TAC TCT 4230

E T K A F E M A G P S K G K E S * 1427
GAA ACT AAG GCC TTT GAG ATG GCT GGA CCC TCC AAA GGT AAA GAA AGC TAG 4281

ATACCCCTCCTTGGAGTTGCAAGTATTCTTTCAAGGTTGGAAGAGGGATTTTGAAGAGGTATCTCTCCAAGCAAGAATGA
CTTGTTTTTCCATAAGGGACATGAGCATTCTTACTAGGCTTGAAGAGCTGACATGATGAGCATTATTGTATGTTGTAT

FIGURE 24D

ATACATTTGTGATAGAGGGCTAGAGTTTGACCTAGAGAGAGTTTAAAGGAAGTGAAATATTTAATTCAGAACCAAATGCT
TTTGTA AAAACTTTTTGGATTTTGTA AAAAGCATTTCATTCCTCTAGAAATTC AAGTATTTTCAAGGGGAGTCATTTGAG
ATATATTTATTTTACTAGGAGATCTTATATTC TAGGAAATGCTTTAAATGGTCAGGCTCCARTCGGAATTTTTTTAAG
AAAAAGTAGTTTTTAATACATTGGTTAGGACTCAGAGGAAATACGGAAAAACATTGTAGATGGTTAATTTACAGATA
AAATCCCAAGAGCCTTTTAAACAACAAGGTACCTAAAAATAGGGTATAATTATACTGCTTAAATACAGTTAGTGCCTAT
TAATAGCTTTTTATTTCTATGGGAAGATGCTTTTGGTCTTCTGGCTGAAGATGTAGGCATACCTCTCACTCATTTCAA
TGTTTTCTTGAGGTGGAGCCTTCATTGGAAAGGGGAAAGAGGGATTCTAGGGTTTCATCAGGGACCAGGAATGCATTCC
TCTGTCAGGTTCCAATCAAGAGAAGACCTTTTATGAGATCTGCCTCTGTATAGATGTTGTCAGTTAGGAACTGAAGCC
ATAGGTCAGGCAGACATCAGCTCAGCCTGTGGCCCATTTGGGTGATTTCTGTATTTTAAACTGACAGTAGCCTGATCA
AAGTGATACAATCAATTTCAAAACAATCTTCCAGAGACCACTTGAAGGTTTCATAGTTTTTACAATACCTGAGACTTTT
CAGGTGTTGGAGCCTCTAAAAATATGAGATATAAACAGAACTAATACAAGTTGTTCTCTGGAGGTTTCTATGAGGTTCT
TAGAAAAATTTGGTTTTAAATCATTTGAGGACAGGAATGTCTATAGCAAGTTTACTCCTATTGCGAATCATGTATGCT
GGCTTTAGTTGTAACAAACGATTTTATTCTAAGTAAGGCCAGGTGCTACTATAAAATCATATATTCCTGTTGAAGTTCT
TTTGAAGTCTATCTCTATTTATTATATTTGAAAGTTGTCAGCCACCAGTCATCCAGAATTTCTTCTCTGAATCTCCAT
GCTCATATGCAATGTCTACATCAAGGTCTTCTTAATGACTATTATTCTCAGGGTTTAGTTTTCTACCTTCTGCCTACTA
TTTTGGTCTGACATTTTGTAGCCTTCTGTTATTATTGGAAATAGTCTCTTACATAAGCTGATTTGAGAACTTTCAAA
ATCTCACATAGCTAATGGAAGTTGCTTTCTGCTTTCTTATGACTGTTTTTATAAATAAACTGTTTCATAAATAAAAAA
AAAAAAAAGGGCGGCCGC

FIGURE 24E

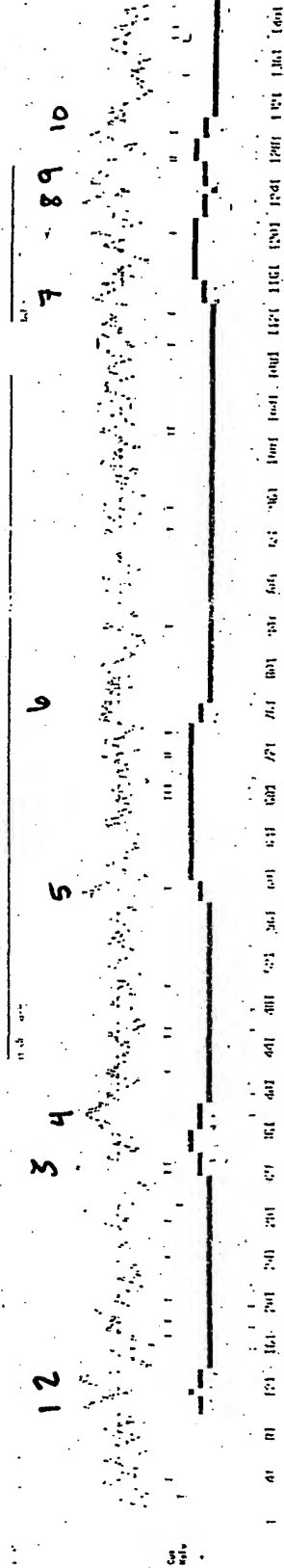


FIGURE 25

Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)
 Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

Query: 67102

Model	Description	Score	E-value	N
Hydrolase 1	haloacid dehalogenase-like hydrolase	1.5	0.17	
DUF6 1	Integral membrane protein DUF6	-24.6	9.4	

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
Hydrolase	1/1	432	1077	1	184	1.5	0.17
DUF6	1/1	1127	1271	1	126	-24.6	9.4

```
*->ikavvFDkDGTLTdgkeppiaeaivealrelgl...apleevekl1
  i  ++ Dk+GTLt+  + + ++ v+ ++  + ++ ++ + + e+
```

```
grgl.g..erilleggltaell.....
+ +   +++ ++l+ +++ ++, + ++ +++++ ++++++++ +++ + +++
```

.....

++ ++ +++++ +++ ++ ++?+ ++ + +++ + +++ ++

.....
+ + + + + + + + + + + + + + + + + + + +

.....

+ + ++++++ ++++++++ ++ ++ + ++ ++ ++ ++ ++ +

.....

+++ ++++ ++++++ ++++ + ++++ ++ ++ + + ++

.....
+++ + ++++++ + ++ + + + ++ ++

.....

+ + + + + + + + + + + + + + + + + + + + + + +

.....

++ ++ + + ++ ++ ++++ + ++++ ++++ ++++++ +

FIGURE 26A

```

.....ld.evlglial.dklypgarealkaLkerGikvailTngdr.na
+. + *+ +lg+ +d l +g+e. +aL+++Gik+++lT++ +a+
67102 878 samrleNKlTLGATGieDRLOEGVPESIEALHKAGIKIWMLTGDKQETA 927
ealle...algla.lfdaiydsdevggvgpvvvgKPkpeifllalerlgv
+ ++ ++l+ ++ + ++ + + + +g + i+++++ +
67102 928 VNIAYackLLEPDdKLFILNTQSKDA-----CGMLMSTILKELOKKTQA 971
kpeevg.....
pe+v+ +++ +++++ + + +++++ + + + + +
67102 972 LPEQVSlsedllqppvprdsgragliitgktlefalqeslqkqflelts 1021
.....p.kvlmvGDginDapalaaAGvgv
+ + ++ +++++ + + + +l++GDg nD+ ++ A++g+
67102 1022 wcqavvccratplqksevklvrshlQvMTLAIGDGANDVSMIQVADIGI 1071
amngng<-*
67102 1072 GVSGQE 1077
DUF6: domain 1 of 1, from 1127 to 1271: score -24.6, E = 9.4
*->fiWalytvfsskille..splttawrfliagilllillflkkgppl
+++ +++++ + + + + + + +l+ + + + + +
67102 1127 LLEWYQFFCGFSGTSmtDYWVLIFFNLLFTS--APPVIYGVLKDV 1170
.....lallslkilallylgilgtalgylyf
+ ++ + ++ +++++ + + + l +y ++++++y+ y
67102 1171 saetlmqlpelyrsgqkseaylpHTFWITL-LDAFYQSLVCFVPYFTYQ 1219
yalkyvsaskasvlsslsPvftlilsvllLgEkltlkqllGivlillGvl
+ ++ + + + +f+++l + + +lt+++ll i+ +l+ +
67102 1220 --GSDTDIFAFGNPLNTAALFIVLLHLVIESKSLTWIHLVIGSILSYF 1267
lisl<-*
l +
67102 1268 LFAI 1271

```

FIGURE 26B

USTAL W (1.74) multiple sequence alignment

```

h67102FL  ---MTEALQWARYHWRRLIRGATRDDDSGPYNYSSLLA-CGRKSSQIPKLSGRHRIVVP
ouseAT5A  MERELPAAEESASSGWRRPRR--RRWEGRTTRTVRSNLLPPLGTEDSTIGAPKGERLLMRG
          .. * : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

          TM1
h67102FL  HIQPFKDEYEKFSGAYVNNRIRTTKYTLLNFVPRNLFQFHRAASLYFLFLVVLNFWPLV
ouseAT5A  CIQHLAD-----NRLKTKYTLLSFLPKNLFQFHRLANVFVFIALINFWPAV
          ** : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

          TM2
h67102FL  EAFQKEITMLPLVVLTIIAIKDGLFEDYRKYKIDKQINNLTQVYSRKEKKYIDRCWKDV
ouseAT5A  NAFQPLGALAPVLFILAVTAIKDLWEDYSRHRSDHEINHLGCLVFSREEKKYVNRWKEI
          : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

          Phospholipid Transport
h67102FL  TVGDFIRLSCNEVI PADMVLLFSTD PDGICHIE TSGLDGESNLKQRQVVRGYAEQDSEVD
ouseAT5A  RVGDFVRLCCNEI I PADILLSSSDPDGLCHIE TANI DGETNLKRRQVVRGFSELVSEFN
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *

h67102FL  PEKFSSRIECESPNNDLSRFRGFLEHSNKERVGLSKENLLLRGCTIRNTEAVVGIVVYAG
ouseAT5A  PLTFTSVIECEKPNNDLSRFRGYIMHSNGEKAGLHKENLLLRGCTIRNTEAVAGIVVIYAG
          * : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

          TM3
h67102FL  HETKAMLNNSGPRYKRSKLERRANTDVLWCVMLLVIMCLTGAVGHGIWISRY-EKMHFFN
ouseAT5A  HETKALLNNSGPRYKRSQLERQMNCLVLWCVLLLVCI SLFSAVGHGLWVRRYQEKKALFD
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *

          TM4
h67102FL  VPEPDGHIISHLLAGFYMFWTMIILLQVLIPISLYVSIEIVKLGQIYFIQSDVDFYNEKM
ouseAT5A  VPESDGSLSLPATAAVYSFFETMIIVLQVLIPISLYVSIEIVKVCQVYFINQDIELYDEET
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *

          Phospholipid Transport > Phosphorylation Site
h67102FL  DSIVQCRALNIAEDLGQIQYIFSDKTGTLTENKMFRRCSVAGFDYCHEENARRLESYQE
ouseAT5A  DSQLOCRALNITEDLGQIKYIFSDKTGTLTENKMFRRCTVSGIEYSHDANAQRLARYQE
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *

bh67102FL  AVSEDEDFIDTVSGSLSNMAKPRAPSCRTVHNGPLGNKPSNHLGSSFTLGS-GEGASEV
ouseAT5A  ADSEEEVVSKV-GTIS---HRGSTG--SHQS-IWMTHKTQSIKSHRRTGSRAEAKRAS
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *

bh67102FL  PHSRQAAFSSPIETDVVPDTRLDDKFSOITPRLFMPLDETIQNPPM---ETLYIIDFFI
ouseAT5A  MLSKHTAFSSPMEKDITPDPKLLKLVSECD-R-FLAIARHQEHPLAHLSPELSDVDFDFI
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *

          TM5
h67102FL  ALAICNTVVVSAIPNQPRQKIRHPSLGGLPIKSLEEI KSLFQRWSVRRSSSPSLNSGKEPS
ouseAT5A  ALTICNTVVVTSPDQPRQKVRVRFELKSPVKTIED---FLRRFTPSRLASGCSSIGNLST
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *

bh67102FL  S-GVPNAFVSRPLPLFSRMKPASPVVEEVSVQCESPQCSSSSACCTETEKQHG DAGLLNGK
ouseAT5A  SKSSHKSGSAFLPSLSQDSMLLGLBEKLGQT--APSIASNGYASQAGQEE SWASDCTT--
          * : * : * * * * * * * * * * * * * * * * * * * * * *

          TM6
bh67102FL  AESLPG-Q-PLACN-LCYEAESPDEAALVYAARAYQCTLRSTPEQVMVDFAALGPLTFQ
ouseAT5A  DQKCPGEQREQQEGELRYEAESPDEAALVYAARAYNCALVDRLHDQVSVELPHLGRLTFE
          : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

bh67102FL  LLHILPFDsvrkrmsvvvrhplsnqvvytkgadsvimellsvaspgdaslekqqmivre
ouseAT5A  LLHTLGFDsirkrmsvvirhpltd einvytkgadsvmdlllpcssddarg-rhqqkirs
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *

bh67102FL  KTQKHLDDYAKQGLRTLCAIAKKVMSDTEYAEWLRNHFLAETSIDNREELLLESAMRLENK

```

FIGURE 27A

[illegible]

FIGURE 27B

Input file Fbh44181pat.seq; Output File Fbh44181pat.tra
Sequence length 7221

GCCGCGGGATGGGAACGCGGCGCGGGGAGTGAGGCAGTGCGCGCGCGCGGTAAGCGGAACTTCGGCCCCGAGGGGCTC

GCCCGCTCCCGCCTCTGTCTTGTCTGGCCTCCACCTGCAGCCCCGCGGCCCCCGCGCCCCGCGGGACCCGGACGGCGACG

M	W	R	W	I	R	Q	Q	L	G	F	D	P	P	H	Q	S	17			
ACGGGGGA	ATG	TGG	CGC	TGG	ATC	CGG	CAG	CAG	CTG	GGT	TTT	GAC	CCA	CCA	CAT	CAG	AGT	51		
D	T	R	T	I	Y	V	A	H	R	F	P	Q	N	G	L	Y	T	P	Q	37
GAC	ACA	AGA	ACC	ATC	TAC	GTA	GCC	CAC	AGG	TTT	CCT	CAG	AAT	GGC	CTT	TAC	ACA	CCT	CAG	111
K	F	I	D	N	R	I	I	S	S	K	Y	T	V	W	N	F	V	P	K	57
AAA	TTT	ATA	GAT	AAC	AGG	ATC	ATT	TCA	TCT	AAG	TAC	ACT	GTG	TGG	AAT	TTT	GTT	CCA	AAA	171
N	L	F	E	Q	F	R	R	V	A	N	F	Y	F	L	I	I	F	L	V	77
AAT	TTA	TTT	GAA	CAG	TTC	AGA	AGA	GTG	GCA	AAC	TTT	TAT	TTT	CTT	ATT	ATA	TTT	TTG	GTT	231
Q	L	M	I	D	T	P	T	S	P	V	T	S	G	L	P	L	F	F	V	97
CAG	CTT	ATG	ATT	GAT	ACA	CCT	ACC	AGT	CCA	GTT	ACC	AGT	GGA	CTT	CCA	TTA	TTC	TTT	GTG	291
I	T	V	T	A	I	K	Q	G	Y	E	D	W	L	R	H	N	S	D	N	117
ATA	ACA	GTA	ACT	GCC	ATA	AAG	CAG	GGA	TAT	GAA	GAT	TGG	TTA	CGG	CAT	AAC	TCA	GAT	AAT	351
E	V	N	G	A	P	V	Y	V	V	R	S	G	G	L	V	K	T	R	S	137
GAA	GTA	AAT	GGA	GCT	CCT	GTT	TAT	GTT	GTT	CGA	AGT	GGT	GGC	CTT	GTA	AAA	ACT	AGA	TCA	411
K	N	I	R	V	G	D	I	V	R	I	A	K	D	E	I	F	P	A	D	157
AAA	AAC	ATT	CGG	GTG	GGT	GAT	ATT	GTT	CGA	ATA	GCC	AAA	GAT	GAA	ATT	TTT	CCT	GCA	GAC	471
L	V	L	L	S	S	D	R	L	D	G	S	C	H	V	T	T	A	S	L	177
TTG	GTG	CTT	CTG	TCC	TCA	GAT	CGA	CTG	GAT	GGT	TCC	TGT	CAC	GTT	ACA	ACT	GCT	AGT	TTG	531
D	G	E	T	N	L	K	T	H	V	A	V	P	E	T	A	L	L	Q	T	197
GAC	GGA	GAA	ACT	AAC	CTG	AAG	ACA	CAT	GTG	GCA	GTT	CCA	GAA	ACA	GCA	TTA	TTA	CAA	ACA	591
V	A	N	L	D	T	L	V	A	V	I	E	C	Q	Q	P	E	A	D	L	217
GTT	GCC	AAT	TTG	GAC	ACT	CTA	GTA	GCT	GTA	ATA	GAA	TGC	CAG	CAA	CCA	GAA	GCA	GAC	TTA	651
Y	R	F	M	G	R	M	I	I	T	Q	Q	M	E	E	I	V	R	P	L	237
TAC	AGA	TTC	ATG	GGA	CGA	ATG	ATC	ATA	ACC	CAA	CAA	ATG	GAA	GAA	ATT	GTA	AGA	CCT	CTG	711
G	P	E	S	L	L	L	R	G	A	R	L	K	N	T	K	E	I	F	G	257
GGG	CCG	GAG	AGT	CTC	CTG	CTT	CGT	GGA	GCC	AGA	TTA	AAA	AAC	ACA	AAA	GAA	ATT	TTT	GGT	771
V	A	V	Y	T	G	M	E	T	K	M	A	L	N	Y	K	S	K	S	Q	277
GTT	GCG	GTA	TAC	ACT	GGA	ATG	GAA	ACT	AAG	ATG	GCA	TTA	AAT	TAC	AAG	AGC	AAA	TCA	CAG	831
K	R	S	A	V	E	K	S	M	N	T	F	L	I	I	Y	L	V	I	L	297
AAA	CGA	TCT	GCA	GTA	GAA	AAG	TCA	ATG	AAT	ACA	TTT	TTG	ATA	ATT	TAT	CTA	GTA	ATT	CTT	891
I	S	E	A	V	I	S	T	I	L	K	Y	T	W	Q	A	E	E	K	W	317
ATA	TCT	GAA	GCT	GTC	ATC	AGC	ACT	ATC	TTG	AAG	TAT	ACA	TGG	CAA	GCT	GAA	GAA	AAA	TGG	951
D	E	P	W	Y	N	Q	K	T	E	H	Q	R	N	S	S	K	I	L	R	337
GAT	GAA	CCT	TGG	TAT	AAC	CAA	AAA	ACA	GAA	CAT	CAA	AGA	AAT	AGC	AGT	AAG	ATT	CTG	AGA	1011

FIGURE 28A

F	I	S	D	F	L	A	F	L	V	L	Y	N	F	I	I	P	I	S	L	357
TTT	ATT	TCA	GAC	TTC	CTT	GCT	TTT	TTG	GTT	CTC	TAC	AAT	TTC	ATC	ATT	CCA	ATT	TCA	TTA	1071
Y	V	T	V	E	M	Q	K	F	L	G	S	F	F	I	G	W	D	L	D	377
TAT	GTG	ACA	GTC	GAA	ATG	CAG	AAA	TTT	CTT	GGA	TCA	TTT	TTT	ATT	GGC	TGG	GAT	CTT	GAT	1131
L	Y	H	E	E	S	D	Q	K	A	Q	V	N	T	S	D	L	N	E	E	397
CTG	TAT	CAT	GAA	GAA	TCA	GAT	CAG	AAA	GCT	CAA	GTC	AAT	ACT	TCC	GAT	CTG	AAT	GAA	GAG	1191
L	G	Q	V	E	Y	V	F	T	D	K	T	G	T	L	T	E	N	E	M	417
CTT	GGA	CAG	GTA	GAG	TAC	GTG	TTT	ACA	GAT	AAA	ACT	GGT	ACA	CTG	ACA	GAA	AAT	GAG	ATG	1251
Q	F	R	E	C	S	I	N	G	M	K	Y	Q	E	I	N	G	R	L	V	437
CAG	TTT	CGG	GAA	TGT	TCA	ATT	AAT	GGC	ATG	AAA	TAC	CAA	GAA	ATT	AAT	GGT	AGA	CTT	GTA	1311
P	E	G	P	T	P	D	S	S	E	G	N	L	S	Y	L	S	S	L	S	457
CCC	GAA	GGA	CCA	ACA	CCA	GAC	TCT	TCA	GAA	GGA	AAC	TTA	TCT	TAT	CTT	AGT	AGT	TTA	TCC	1371
H	L	N	N	L	S	H	L	T	T	S	S	S	F	R	T	S	P	E	N	477
CAT	CTT	AAC	AAC	TTA	TCC	CAT	CTT	ACA	ACC	AGT	TCC	TCT	TTC	AGA	ACC	AGT	CCT	GAA	AAT	1431
E	T	E	L	I	K	E	H	D	L	F	F	K	A	V	S	L	C	H	T	497
GAA	ACT	GAA	CTA	ATT	AAA	GAA	CAT	GAT	CTC	TTC	TTT	AAA	GCA	GTC	AGT	CTC	TGT	CAC	ACT	1491
V	Q	I	S	N	V	Q	T	D	C	T	G	D	G	P	W	Q	S	N	L	517
GTA	CAG	ATT	AGC	AAT	GTT	CAA	ACT	GAC	TGC	ACT	GGT	GAT	GGT	CCC	TGG	CAA	TCC	AAC	CTG	1551
A	P	S	Q	L	E	Y	Y	A	S	S	P	D	E	K	A	L	V	E	A	537
GCA	CCA	TCG	CAG	TTG	GAG	TAC	TAT	GCA	TCT	TCA	CCA	GAT	GAA	AAG	GCT	CTA	GTA	GAA	GCT	1611
A	A	R	I	G	I	V	F	I	G	N	S	E	E	T	M	E	V	K	T	557
GCT	GCA	AGG	ATT	GGT	ATT	GTG	TTT	ATT	GGC	AAT	TCT	GAA	GAA	ACT	ATG	GAG	GTT	AAA	ACT	1671
L	G	K	L	E	R	Y	K	L	L	H	I	L	E	F	D	S	D	R	R	577
CTT	GGA	AAA	CTG	GAA	CGG	TAC	AAA	CTG	CTT	CAT	ATT	CTG	GAA	TTT	GAT	TCA	GAT	CGT	AGG	1731
R	M	S	V	I	V	Q	A	P	S	G	E	K	L	L	F	A	K	G	A	597
AGA	ATG	AGT	GTA	ATT	GTT	CAG	GCA	CCT	TCA	GGT	GAG	AAG	TTA	TTA	TTT	GCT	AAA	GGA	GCT	1791
E	S	S	I	L	P	K	C	I	G	G	E	I	E	K	T	R	I	H	V	617
GAG	TCA	TCA	ATT	CTC	CCT	AAA	TGT	ATA	GGT	GGA	GAA	ATA	GAA	AAA	ACC	AGA	ATT	CAT	GTA	1851
D	E	F	A	L	K	G	L	R	T	L	C	I	A	Y	R	K	F	T	S	637
GAT	GAA	TTT	GCT	TTG	AAA	GGG	CTA	AGA	ACT	CTG	TGT	ATA	GCA	TAT	AGA	AAA	TTT	ACA	TCA	1911
K	E	Y	E	E	I	D	K	R	I	F	E	A	R	T	A	L	Q	Q	R	657
AAA	GAG	TAT	GAG	GAA	ATA	GAT	AAA	CGC	ATA	TTT	GAA	GCC	AGG	ACT	GCC	TTG	CAG	CAG	CGG	1971
E	E	K	L	A	A	V	F	Q	F	I	E	K	D	L	I	L	L	G	A	677
GAA	GAG	AAA	TTG	GCA	GCT	GTT	TTC	CAG	TTC	ATA	GAG	AAA	GAC	CTG	ATA	TTA	CTT	GGA	GCC	2031
T	A	V	E	D	R	L	Q	D	K	V	R	E	T	I	E	A	L	R	M	697
ACA	GCA	GTA	GAA	GAC	AGA	CTA	CAA	GAT	AAA	GTT	CGA	GAA	ACT	ATT	GAA	GCA	TTG	AGA	ATG	2091
A	G	I	K	V	W	V	L	T	G	D	K	H	E	T	A	V	S	V	S	717
GCT	GGT	ATC	AAA	GTA	TGG	GTA	CTT	ACT	GGG	GAT	AAA	CAT	GAA	ACA	GCT	GTT	AGT	GTG	AGT	2151

FIGURE 28B

L	S	C	G	H	F	H	R	T	M	N	I	L	E	L	I	N	Q	K	S	737
TTA	TCA	TGT	GGC	CAT	TTT	CAT	AGA	ACC	ATG	AAC	ATC	CTT	GAA	CTT	ATA	AAC	CAG	AAA	TCA	2211
D	S	E	C	A	E	Q	L	R	Q	L	A	R	R	I	T	E	D	H	V	757
GAC	AGC	GAG	TGT	GCT	GAA	CAA	TTG	AGG	CAG	CTT	GCC	AGA	AGA	ATT	ACA	GAG	GAT	CAT	GTG	2271
I	Q	H	G	L	V	V	D	G	T	S	L	S	E	A	L	R	E	H	E	777
ATT	CAG	CAT	GGG	CTG	GTA	GTG	GAT	GGG	ACC	AGC	CTA	TCT	CTT	GCA	CTC	AGG	GAG	CAT	GAA	2331
K	L	F	M	E	V	C	R	N	C	S	A	V	L	C	C	R	M	A	P	797
AAA	CTA	TTT	ATG	GAA	GTT	TGC	AGA	AAT	TGT	TCA	GCT	GTA	TTA	TGC	TGT	CGT	ATG	GCT	CCA	2391
L	Q	K	A	K	V	I	R	L	I	K	I	S	P	E	K	P	I	T	L	817
CTG	CAG	AAA	GCA	AAA	GTA	ATA	AGA	CTA	ATA	AAA	ATA	TCA	CCT	GAG	AAA	CCT	ATA	ACA	TTG	2451
A	V	G	D	G	A	N	D	V	S	M	I	Q	E	A	H	V	G	I	G	837
GCT	GTT	GGT	GAT	GGT	GCT	AAT	GAC	GTA	AGC	ATG	ATA	CAA	GAA	GCC	CAT	GTT	GGC	ATA	GGA	2511
I	M	G	K	E	G	R	Q	A	A	R	N	S	D	Y	A	I	A	R	F	857
ATC	ATG	GGT	AAA	GAA	GGA	AGA	CAG	GCT	GCA	AGA	AAC	AGT	GAC	TAT	GCA	ATA	GCC	AGA	TTT	2571
K	F	L	S	K	L	L	F	V	H	G	H	F	Y	Y	I	R	I	A	T	877
AAG	TTC	CTC	TCC	AAA	TTG	CTT	TTT	GTT	CAT	GGT	CAT	TTT	TAT	TAT	ATT	AGA	ATA	GCT	ACC	2631
L	V	Q	Y	F	F	Y	K	N	V	C	F	I	T	P	Q	F	L	Y	Q	897
CTT	GTA	CAG	TAT	TTT	TTT	TAT	AAG	AAT	GTG	TGC	TTT	ATC	ACA	CCC	CAG	TTT	TTA	TAT	CAG	2691
F	Y	C	L	F	S	Q	Q	T	L	Y	D	S	V	Y	L	T	L	Y	N	917
TTC	TAC	TGT	TTG	TTT	TCT	CAG	CAA	ACA	TTG	TAT	GAC	AGC	GTG	TAC	CTG	ACT	TTA	TAC	AAT	2751
I	C	F	T	S	L	P	I	L	I	Y	S	L	L	E	Q	H	V	D	P	937
ATT	TGT	TTT	ACT	TCC	CTA	CCT	ATT	CTG	ATA	TAT	AGT	CTT	TTG	GAA	CAG	CAT	GTA	GAC	CCT	2811
H	V	L	Q	N	K	P	T	L	Y	R	D	I	S	K	N	R	L	L	S	957
CAT	GTG	TTA	CAA	AAT	AAG	CCC	ACC	CTT	TAT	CGA	GAC	ATT	AGT	AAA	AAC	CGC	CTC	TTA	AGT	2871
I	K	T	F	L	Y	W	T	I	L	G	F	S	H	A	F	I	F	F	F	977
ATT	AAA	ACA	TTT	CTT	TAT	TGG	ACC	ATC	CTG	GGC	TTC	AGT	CAT	GCC	TTT	ATT	TTC	TTT	TTT	2931
G	S	Y	L	L	I	G	K	D	T	S	L	L	G	N	G	Q	M	F	G	997
GGA	TCC	TAT	TTA	CTA	ATA	GGG	AAA	GAT	ACA	TCT	CTG	CTT	GGA	AAT	GGC	CAG	ATG	TTT	GGA	2991
N	W	T	F	G	T	L	V	F	T	V	M	V	I	T	V	T	V	K	M	1017
AAC	TGG	ACA	TTT	GGC	ACT	TTG	GTC	TTC	ACA	GTC	ATG	GTT	ATT	ACA	GTC	ACA	GTA	AAG	ATG	3051
A	L	E	T	H	F	W	T	W	I	N	H	L	V	T	W	G	S	I	I	1037
GCT	CTG	GAA	ACT	CAT	TTT	TGG	ACT	TGG	ATC	AAC	CAT	CTC	GTT	ACC	TGG	GGA	TCT	ATT	ATA	3111
F	Y	F	V	F	S	L	F	Y	G	G	I	L	W	P	F	L	G	S	Q	1057
TTT	TAT	TTT	GTA	TTT	TCC	TTG	TTT	TAT	GGA	GGG	ATT	CTC	TGG	CCA	TTT	TTG	GGC	TCC	CAG	3171
N	M	Y	F	V	F	I	Q	L	L	S	S	G	S	A	W	F	A	I	I	1077
AAT	ATG	TAT	TTT	GTG	TTT	ATT	CAG	CTC	CTG	TCA	AGT	GGT	TCT	GCT	TGG	TTT	GCC	ATA	ATC	3231
L	M	V	V	T	C	L	F	L	D	I	I	K	K	V	F	D	R	H	L	1097
CTC	ATG	GTT	GTT	ACA	TGT	CTA	TTT	CTT	GAT	ATC	ATA	AAG	AAG	GTC	TTT	GAC	CGA	CAC	CTC	3291

FIGURE 28C

H P T S T E K A Q L T E T N A G I K C L 1117
CAC CCT ACA AGT ACT GAA AAG GCA CAG CTT ACT GAA ACA AAT GCA GGT ATG AAG TGC TTG 3351

D S M C C F P E G E A A C A S V G R M L 1137
GAC TCC ATG TGC TGT TTC CCG GAA GGA GAA GCA GCG TGT GCA TCT GTT GGA AGA ATG CTG 3411

E R V I G R C S P T H I S S S W S A S D 1157
GAA CGA GTT ATA GGA AGA TGT AGT CCA ACC CAC ATC AGC AGT TCA TGG AGT GCA TCG GAT 3471

P F Y T N D R S I L T L S T M D S S T C 1177
CCT TTC TAT ACC AAC GAC AGG AGC ATC TTG ACT CTC TCC ACA ATG GAC TCA TCT ACT TGT 3531

TAA 1178
3534

AGGGGCAGTAGTACTTTGTGGGAGCCAGTTCACCTCCTTTCTTAAATTCAGTGTGATCACCTGTAAATGGCCACACT
AGCTCTGAAATTAATTTCCAAAATCTTTGTAGTAGTTCATACCCACTCAGAGTTATAATGGCAAACAAACAGAAAGCAT
TAGTACAAGCCCCTCCCAACACCCTTAATTTGAATCTGAACATGTTAAAATTTGAGAATAAAGAGACATTTTTCATCTC
TTTGTCTGGTTTGTCCCTTGTGCTTATGGGACTCCTAATGGCATTTCAGTCTGTTGCTGAGGCCATTATATTTTAATAT
AAATGTAGAAAAAAGAGAGAAATCTTAGTAAAGAGTATTTTGTAGTATTAGCTTGATTATTGACTCTTCTATTTAAATC
TGCTTCTGTAAATTATGCTGAAAGTTTGCCTTGAGAACTCTATTTTTTTTATTAGAGTTATATTTAAAGCTTTTTCATGGG
AAAAGTTAATGTGAATACTGAGGAATTTTGGTCCCTCAGTGACCTGTGTTGTAAATTCATTAATGCATTCTGAGTTCAC
AGAGCAAATTAGGAGAATCATTTCCAACCATTATTTACTGCAGTATGGGGAGTAAATTTATACCAATTCCTCTAACTGT
ACTGTAACACAGCCTGTAAAGTTAGCCATATAAATGCAAGGGTATATCATATATACAAATCAGGAATCAGGTCCGTTCA
CCGAACCTCAAATTGATGTTTACTAATATTTTGTGACAGAGTATAAAGACCCTATAGTGGGTAAATTAGATACTATTA
GCATATTATTAATTTAATGTCTTTATCATTTGGATCTTTTGCATGCTTTAATCTGGTTAACATATTTAAATTTGCTTTTT
TTCTCTTTACCTGAAGGCTCTGTGTATAGTATTTTCATGACATCGTTGTACAGTTTAACTATCAATAAAAAAGTTTGGACA
GTATTTAAATATTGCAAATATGTTTAAATTATACAAATCAGAATAGTATGGGTAATTTAAATGAATACAAAAAGAAGAGCC
TCTTTCTGCAGCCGACTTAGACATGCTCTTCCCTTTCTATAAGCTAGATTTTAGAATAAAGGGTTTCAGTTAATAATCT
TATTTTCAGGTTATGTCATCTAACTTATAGCAAATACCACAATACAGTGAGTTCTGCCAGTGTCCAGTACAAGGCAT
ATTTTCAGGTGTGGCTGTGGAATGTAAAAATGCTCAACTTGTATCAGGTAATGTTAGCAATAAATTTAAATGCTAAGAATG
ATTAATCGGGTACATGTTACTGTAATTAACCTATTGCACCTCAAACCTAACTTCCATCCTGAATTTATCAAGTAGTTC
AGTATTGTCATTTGTTTTTGTGTTTTATTGAAAAGTAATGTTGCTTTAAGATTTAGAAGTGATTATTAGCTTGAGAACTAT
TACCCAGCTCTAAGCAAATAATGATTGTATACATATTAAGATAATGGTTAAATGCGGTTTTACCAAGTTTCCCTTGAA
AATGTAATTCCTTTATGGAGATTTATTGTGCAGCCCTAAGCTTCCTTCCCATTTTCATGAATATAAGGCTTCTAGAATTG
GACTGGCAGGGGAAAGAATGGTAGAGACAGAAATTAAGACTTTATCCTTGTTTGCTTGTAACCTATTATTTTCTTGCTA
ATGTAACATTTGTCTGTTCCAGTGATGTAAGGATATTAAGTTATTAAGCTAAATATTAATTTTCAAAAATAGTCCTTCT

FIGURE 28D

TTAACTTAGATATTTTCATAGCTGGATTAGGAAGATCTGTTATTCTGGAAGTACTAAAAAGAATAATACAACGTACAAT
GTCTGCATTCACTAATTCATGTTCCAGAAGAGGAAATAATGAAGATATACTCAGTAGAGTACTAGGTGGGAGGATATGG
AAATTTGCTCATAAAATCTCTTATAAAACGTGCATATAACAAAATGACACCCAGTAGGCCTGCATTACATTTACATGAC
CGTGTTTATTTGCCATCAAATAAACTGAGTACTGACACCAGACAAAGACTCCAAAGTCATAAAATAGCCTATGACCAAC
TGCAGCAAGACAGGAGGTCAGCTCGCCTATAATGGTGCTTAAAGTGTGATTGATGTAATTTTCTGTACTCACCATTTGA
AGTTAGTTAAGGAGAAGCTTTATTTTTTTAAAAAAAGTAAATGGCAACCACCTAGTGTGCTCATCTGAAGCTTTACTCCA
AATCCACTCCGTTTTTAAAGCAAAATATCTTGTGATTTTAAGAAAAGAGTTTTCTATTTATTTAAGAAAGTAACAATG
CAGTCTGCAAGCTTTCAGTAGTTTTCTAGTGCTATATTCATCCTGTAAACCTTTACTACGTAACCAGTAATCACAAGG
AAAGTGTCCCTTTGCATATTTCTTTAAAAATCTTTCTTTGGAAAGTATGATGTTGATAATTAACCTTACCTTTATCTGC
CAAAACCAGAGCAAAATGCTAAATACGTTATTGCTAATCAGTGGTCTCAAATCGATTGCTCCCTTTGCCCTCGTCTGA
GGGCTGTAAGCCTGAAGATAGTGGAAGCACCAAGTCAGTTTCCAAAATGCCCCTCAGCTGCTTTAAGTGACTCAGCA
CCCTGCCTCAGCTTCAGCAGGCGTAGGCTCACCTGGGCGGAGCAAAGTATGGGCCAGGGAGAAGTACAGCTACGAAGA
CCTGCTGTGAGTTGAGAAAAGGGGAGAAATTTATGGTCTGAATTTTCTAACTGTCTCTTTCTTTGGGTCTAAAGCTCAT
AATACACAAAGGCTTCCAGACCTGAGCCACACCCAGGCCCTATCCTGAACAGGAGACTAAACAGAGGCAAAATCAACCTT
AGGAAATACTTGCATTCTGCCCTACGGTTAGTACCAGGACTGAGGTCATTTCTACTGGAAAAGATTGTGAGATTGAACT
TATCTGATCGCTTGAGACTCCTAATAGGCAGGAGTCAAGGCCACTAGAAAATTGACAGTTAAGAGCCAAAAGTTTTTAA
AATATGCTACTCTGAAAAATCTCGTGAAGGCTGTAGGAAAAGGGAGAATCTTCCATGTTGGTGTTTTTCTGTAAAGAT
CAGTTTGGGGTATGATATAAGCAGGTATTAATAAAAAATAACACACCAAAGAGTTACGTAAAACATGTTTTATTAATTTT
GGTCCCCACGTACAGACATTTTATTTCTATTTTGAAATGAGTTATCTATTTTCATAAAAGTAAAACACTATTAAAGTGC
TGTTTTATGTGAAATAACTTGAATGTTGTTCCATAAAAAATAGATCATAACTCATGATATGTTTGTAATCATGGTAAT
TTAGATTTTTATGAGGAATGAGTATCTGGAAATATTGTAGCAATACTTGGTTTTAAAAATTTGGACCTGAGACACTGTGG
CTGTCTAATGTAATCCTTTAAAAATCTCTGCATTGTGAGTAAATGTAGTATATTATTGTACAGCTACTCATAATTTTTT
TAAAGTTTATGAAGTTATATTTATCAAATAAAAACTTTCCTATAT

FIGURE 28E

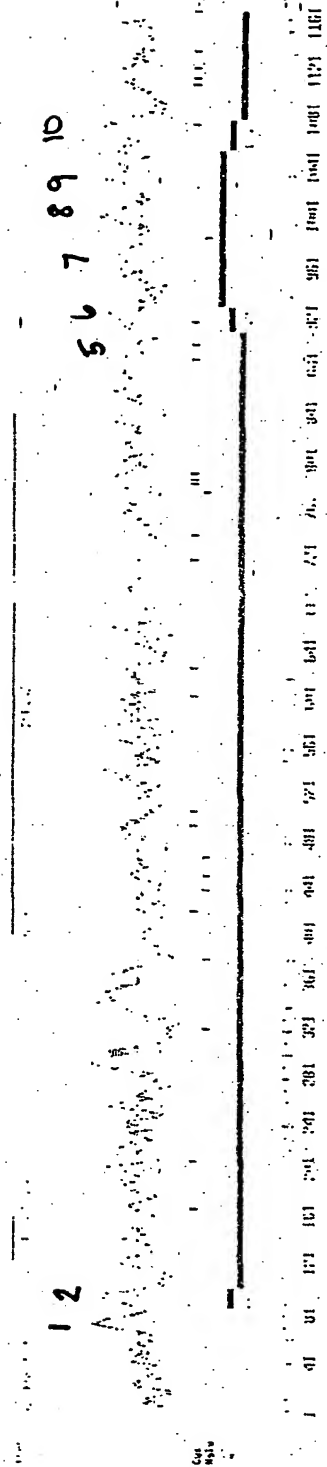


FIGURE 29

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)
 Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.15759.seq

Query: 44181

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Hydrolase	haloacid dehalogenase-like hydrolase	42.8	8e-09	
1				
E1-E2_ATPase	E1-E2' ATPase	8.6	0.13	
1				
DUF132	Protein of unknown function DUF132	-72.9	9.4	
1				

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
E1-E2_ATPase	1/1	126	164	37	75	8.6	0.13
DUF132	1/1	579	719	1	160	-72.9	9.4
Hydrolase	1/1	401	842	1	184	42.8	8e-09

Alignments of top-scoring domains:

E1-E2_ATPase: domain 1 of 1, from 126 to 164: score 8.6, E = 0.13

```

      *->VlRdGkeeeipaeelvpGDIVevkpgDrVPADgrvvege<-*
      V+R G++++ ++ +++GDIV+++ ++ PAD+++++++
44181 126 VVRSGGLVKTRSKNIRVGDIVRIAKDEIFPADLVLLSSD 164

```

DUF132: domain 1 of 1, from 579 to 719: score -72.9, E = 9.4

```

      *->MeeiklkVVIDTsvliaA....LispkGlaflklellfeeKleN..
      M          V++ A+++++L+ kG  +l  +++ +e ++
44181 579 MS-----VIVQApsegekLLFAKGAESSILPKCIGGEIEKtr 614

```

```

      .YtSdeiLeEyifkillpKLecklpvEvslkkvl.vvlvskSkvinPRSF
      ++ + L+ + + l + +k+ + E  +++++ + + +
44181 615 iHVEDEFALKGL--RTLCAIYRKETSKE--YEEIDkRIFEARTALQQR--- 657

```

```

      KESntkFnvCRDpeDNKFLn...vvYesKAdylITYDkDLLdRLRDENkk
      +k  +  F++++ +  + A  +L d+ R
44181 658 ---EEKLAHV-----FOFIEkdlILLGATA-----VEDRLQDKVRETIEA 694

```

```

      lkledHefkvLTPkEFiesveKkls<-*
      l++ +++++ vLT      v  +ls
44181 695 LRMAGIKVWVLTGDKHETAVSVLS 719

```

Hydrolase: domain 1 of 1, from 401 to 842: score 42.8, E = 8e-09

```

      *->ikavvFDkDGTltdgk.....
      +  v+ Dk+GTLt+ + + ++ + ++ + ++ ++++++
44181 401 VEYVFTDKTGTLTENEmqfreccsingmkyqeingrlvpegptpsse 447

```

```

      .....
      ++ + ++ ++ ++ ++ +++++ ++++++ +++++ + + ++
44181 448 gnlsylsslshlnnlshlttsssftrtspenelikehldffkavslcht 497

```

```

      .....eppiaaeavealrelgl....
      + ++ ++ +++++ ++ + + + p ++avea++++g+ +

```

FIGURE 30A

```

44181 498 vqisnvqtdctgdgppwqsnlapsqleyyaSSPDEKALVEAAARIGIvfig 547
      apleevekllgrgl.g...erilleggltaell.....
      +++e +e   +++ l++ + ++il++   +++++   +++++   ++
44181 548 NSEETMEVKTLGKLeRyklLHILEFDSDRRRMSvivqapsgeklfakga 597
      .....
      +++ ++   +++ +++++ + ++   ++ ++   ++ +++++ ++ +++
44181 598 essilpkciggeiektrihvdefalkglrtlciayrkftskeyeekri 647
      .....ld.evlglial.dklypgarealkalke
      + ++   +++++   + d +lg+ a++d l + +re++aL+
44181 648 feartalqqreeklaavfqfieKdlILLGATAVEDRLQDKVRETIEALRM 697
      rGikvailTngdr.naeallealgla.lfdaivdsdevggvgpvvvgKPk
      +Gikv++lT++ +++++a+   +   +   +   +   +   +   +   +   +   +
44181 698 AGIKVWVLTGDKHeTAVS-----VSLSGCHFHRMTNlL---ELINQKSD 738
      peifllalerlgvkepvevg.....
      e+++++++ ++ e ++   +++   +++++ +   +++++ +   ++
44181 739 SECAEQLRQLARRITE--Dhviqhgglvvdgtslslalreheklfmevcn 786
      .....p.kvlmvGDginDapalaaAGvgv
      +   +   +   ++ +   +   +   +   +   +   +   +   +   +   +   +   +
44181 787 csavlcrcmaplgkakvirlikispeKpITLAVGDGANDVSMIQEAHVGI 836
      amgngg<--*
      +
44181 837 GIMGKE      842

```

FIGURE 30B

FIGURE 31B

Input file Fbh67084FL.seq; Output File Fbh67084FL.tra
Sequence length 4198

GGAGTCGACCCACGCGTCCGCATTGAGACAATGCCTCCACAAATACTTGATGCAAATTCAGTAAGACAGCACTTGTTG

AATCACCATTATAGTTTCTGACAAATTGTTCTCAAAAAGGTACCAGCTGGAGGATGAGTCTGCGCATTGATGAA

M	P	L	M	M	S	E	E	G	F	E	N	E	E	S	D	Y	H	T	L	20
ATG	CCA	CTA	ATG	ATG	TCT	GAA	GAA	GGC	TTT	GAG	AAT	GAG	GAA	AGT	GAT	TAC	CAC	ACC	TTA	60
P	R	A	R	I	M	Q	R	K	R	G	L	E	W	F	V	C	D	G	W	40
CCA	CGA	GCC	AGG	ATA	ATG	CAA	AGG	AAA	AGA	GGA	CTG	GAG	TGG	TTT	GTC	TGT	GAT	GGC	TGG	120
K	F	L	C	T	S	C	C	G	W	L	I	N	I	C	R	R	K	K	E	60
AAG	TTT	CTC	TGT	ACC	AGT	TGC	TGT	GGT	TGG	CTG	ATA	AAT	ATT	TGT	CGA	AGA	AAG	AAA	GAG	180
L	K	A	R	T	V	W	L	G	C	P	E	K	C	E	E	K	H	P	R	80
CTG	AAA	GCT	CGC	ACA	GTA	TGG	CTT	GGA	TGT	CCT	GAA	AAG	TGT	GAA	GAA	AAA	CAT	CCC	AGG	240
N	S	I	K	N	Q	K	Y	N	V	F	T	F	I	P	G	V	L	Y	E	100
AAT	TCT	ATA	AAA	AAT	CAA	AAA	TAC	AAT	GTG	TTT	ACC	TTT	ATA	CCT	GGG	GTT	TTG	TAT	GAA	300
Q	F	K	F	F	L	N	L	Y	F	L	V	I	S	C	S	Q	F	V	P	120
CAA	TTC	AAG	TTT	TTC	TTG	AAT	CTC	TAT	TTT	CTA	GTG	ATA	TCC	TGC	TCA	CAG	TTT	GTA	CCA	360
A	L	K	I	G	Y	L	Y	T	Y	W	A	P	L	G	F	V	L	A	V	140
GCA	TTG	AAA	ATA	GGC	TAT	CTC	TAC	ACC	TAC	TGG	GCT	CCT	CTG	GGA	TTT	GTC	TTG	GCT	GTT	420
T	M	T	R	E	A	I	D	E	F	R	R	F	Q	R	D	K	E	V	N	160
ACT	ATG	ACA	CGG	GAA	GCA	ATT	GAT	GAA	TTT	CGG	CGT	TTT	CAG	CGT	GAC	AAG	GAA	GTG	AAT	480
S	Q	L	Y	S	K	L	T	V	R	G	K	V	Q	V	K	S	S	D	I	180
TCA	CAA	CTA	TAT	AGC	AAG	CTT	ACA	GTA	AGA	GGT	AAA	GTG	CAA	GTT	AAG	AGT	TCA	GAC	ATA	540
Q	V	G	D	L	I	I	V	E	K	N	Q	R	I	P	S	D	M	V	F	200
CAA	GTT	GGA	GAC	CTC	ATC	ATA	GTG	GAA	AAG	AAT	CAA	AGA	ATT	CCA	TCG	GAC	ATG	GTG	TTT	600
L	R	T	S	E	K	A	G	S	C	F	I	R	T	D	Q	L	D	G	E	220
CTT	AGG	ACT	TCA	GAA	AAA	GCA	GGT	TCG	TGT	TTT	ATT	CGA	ACT	GAT	CAA	CTA	GAT	GGT	GAA	660
T	D	W	K	L	K	V	A	V	S	C	T	Q	Q	L	P	A	L	G	D	240
ACT	GAC	TGG	AAG	CTG	AAG	GTG	GCA	GTG	AGC	TGC	ACG	CAA	CAG	CTG	CCG	GCT	CTG	GGG	GAC	720
L	F	S	I	S	A	Y	V	Y	A	Q	K	P	Q	M	D	I	H	S	F	260
CTT	TTT	TCT	ATC	AGT	GCT	TAT	GTT	TAT	GCT	CAG	AAA	CCA	CAA	ATG	GAC	ATT	CAC	AGT	TTC	780
E	G	T	F	T	R	E	D	S	D	P	P	I	H	E	S	L	S	I	E	280
GAA	GGC	ACA	TTT	ACC	AGG	GAA	GAC	AGT	GAC	CCG	CCC	ATT	CAT	GAA	AGT	CTC	AGC	ATA	GAA	840
N	T	L	W	A	S	T	I	V	A	S	G	T	V	I	G	V	V	I	Y	300
AAT	ACA	TTG	TGG	GCA	AGC	ACC	ATT	GTT	GCA	TCA	GGT	ACT	GTA	ATA	GGT	GTT	GTC	ATT	TAT	900
T	G	K	E	T	R	S	V	M	N	T	S	N	P	K	N	K	V	G	L	320
ACC	GGA	AAA	GAG	ACT	CGA	AGT	GTA	ATG	AAC	ACA	TCC	AAT	CCA	AAA	AAT	AAG	GTT	GGT	TTG	960
L	D	L	E	L	N	R	L	T	K	A	L	F	L	A	L	V	A	L	S	340
TTG	GAC	CTT	GAA	CTC	AAT	CGG	CTG	ACG	AAA	GCG	CTA	TTT	TTG	GCT	TTA	GTT	GCT	CTT	TCC	1020

FIGURE 32A

I	V	M	V	T	L	Q	G	F	V	G	P	W	Y	R	N	L	F	R	F	360
ATT	GTT	ATG	GTA	ACC	TTA	CAA	GGA	TTT	GTG	GCT	CCA	TGG	TAC	CGC	AAT	CTT	TTT	CGG	TTC	1080
L	L	L	F	S	Y	I	I	P	I	S	L	R	V	N	L	D	M	G	K	380
CTT	CTC	CTC	TTT	TCT	TAC	ATC	ATT	CCC	ATA	AGT	TTG	CGT	GTG	AAC	TTG	GAC	ATG	GGC	AAA	1140
A	V	Y	G	W	M	M	M	K	D	E	N	I	P	G	T	V	V	R	T	400
GCG	GTG	TAT	GGA	TGG	ATG	ATG	ATG	AAA	GAT	GAG	AAC	ATC	CCT	GGC	ACG	GTC	GTT	CGG	ACC	1200
S	T	I	P	E	E	L	G	R	L	V	Y	L	L	T	D	K	T	G	T	420
AGC	ACT	ATC	CCA	GAG	GAA	CTT	GGG	CGC	CTG	GTG	TAT	TTA	TTG	ACA	GAC	AAA	ACA	GGA	ACC	1260
L	T	Q	N	E	M	I	F	K	R	L	H	L	G	T	V	S	Y	G	A	440
CTC	ACC	CAG	AAT	GAA	ATG	ATA	TTT	AAG	CGG	CTG	CAC	CTG	GGC	ACC	GTG	TCC	TAT	GGC	GCC	1320
D	T	M	D	E	I	Q	S	H	V	R	D	S	Y	S	Q	M	Q	S	Q	460
GAC	ACG	ATG	GAT	GAG	ATC	CAG	AGC	CAT	GTC	AGG	GAC	TCC	TAC	TCA	CAG	ATG	CAG	TCT	CAA	1380
A	G	G	N	N	T	G	S	T	P	L	R	K	A	Q	S	S	A	P	K	480
GCT	GGT	GGA	AAC	AAT	ACT	GGT	TCA	ACT	CCA	CTA	AGA	AAA	GCC	CAA	TCT	TCA	GCT	CCC	AAA	1440
V	R	K	S	V	S	S	R	I	H	E	A	V	K	A	I	V	L	C	H	500
GTT	AGG	AAA	AGT	GTC	AGT	AGT	CGA	ATC	CAT	GAA	GCC	GTG	AAA	GCC	ATC	GTG	CTG	TGT	CAC	1500
N	V	T	P	V	Y	E	S	R	A	G	V	T	E	E	T	E	F	A	E	520
AAC	GTG	ACC	CCC	GTG	TAT	GAG	TCT	CGG	GCC	GGC	GTT	ACT	GAG	GAG	ACT	GAG	TTC	GCA	GAG	1560
A	D	Q	D	F	S	D	E	N	R	T	Y	Q	A	S	S	P	D	E	V	540
GCT	GAC	CAA	GAC	TTC	AGT	GAT	GAG	AAT	CGC	ACC	TAC	CAG	GCT	TCC	AGC	CCG	GAT	GAG	GTC	1620
A	L	V	Q	W	T	E	S	V	G	L	T	L	V	S	R	D	L	T	S	560
GCT	CTG	GTG	CAG	TGG	ACA	GAG	AGT	GTG	GGC	CTC	ACG	CTG	GTC	AGC	AGG	GAC	CTC	ACC	TCC	1680
M	Q	L	K	T	P	S	G	Q	V	L	S	F	C	I	L	Q	L	F	P	580
ATG	CAG	CTG	AAG	ACC	CCC	AGT	GGC	CAG	GTC	CTC	AGC	TTC	TGC	ATT	CTG	CAG	CTG	TTT	CCC	1740
F	T	S	E	S	K	R	M	G	V	I	V	R	D	E	S	T	A	E	I	600
TTC	ACC	TCC	GAG	AGC	AAG	CGG	ATG	GGC	GTC	ATC	GTC	AGG	GAT	GAA	TCC	ACG	GCA	GAA	ATC	1800
T	F	Y	M	K	G	A	D	V	A	M	S	P	I	V	Q	Y	N	D	W	620
ACA	TTC	TAC	ATG	AAG	GGC	GCT	GAC	GTG	GCC	ATG	TCT	CCT	ATC	GTG	CAG	TAT	AAT	GAC	TGG	1860
L	E	E	E	C	G	N	M	A	R	E	G	L	R	T	L	V	V	A	K	640
CTG	GAA	GAG	GAG	TGC	GGA	AAC	ATG	GCT	CGC	GAA	GGA	CTG	CGG	ACC	CTC	GTG	GTT	GCA	AAG	1920
K	A	L	T	E	E	Q	Y	Q	D	F	E	S	R	Y	T	Q	A	K	L	660
AAG	GCG	TTG	ACA	GAG	GAG	CAG	TAC	CAG	GAC	TTT	GAG	AGC	CGA	TAC	ACT	CAA	GCC	AAG	CTG	1980
S	M	H	D	R	S	L	K	V	A	A	V	V	E	S	L	E	R	E	M	680
AGC	ATG	CAC	GAC	AGG	TCC	CTC	AAG	GTG	GCC	GCG	GTA	GTC	GAG	AGC	CTG	GAG	AGG	GAG	ATG	2040
E	L	L	C	L	T	G	V	E	D	Q	L	Q	A	D	V	R	P	T	L	700
GAA	CTG	CTG	TGC	CTC	ACC	GGC	GTG	GAG	GAC	CAG	CTG	CAG	GCA	GAC	GTG	CGG	CCC	ACG	CTG	2100
E	M	L	R	N	A	G	I	K	I	W	M	L	T	G	D	K	L	E	T	720
GAG	ATG	CTG	CGC	AAC	GCC	GGG	ATC	AAG	ATA	TGG	ATG	CTA	ACA	GGC	GAT	AAA	CTC	GAG	ACA	2160

FIGURE 32B

A	T	C	I	A	K	S	S	H	L	V	S	R	T	Q	D	I	H	I	F	740
GCT	ACC	TGC	ATT	GCC	AAA	AGT	TCA	CAT	CTC	GTG	TCT	AGA	ACA	CAA	GAT	ATT	CAT	ATT	TTC	2220
R	Q	V	T	S	R	G	E	A	H	L	E	L	N	A	F	R	R	K	H	760
AGA	CAG	GTA	ACC	AGT	CGG	GGA	GAG	GCA	CAT	TTG	GAG	CTG	AAT	GCA	TTT	CGA	AGG	AAG	CAT	2280
D	C	A	L	V	I	S	G	D	S	L	E	V	C	L	K	Y	Y	E	H	780
GAT	TGT	GCA	CTA	GTC	ATA	TCT	GGG	GAC	TCT	CTG	GAG	GTT	TGT	CTA	AAG	TAC	TAC	GAG	CAT	2340
E	F	V	E	L	A	C	Q	C	P	A	V	V	C	C	R	C	S	P	T	800
GAA	TTT	GTG	GAG	CTG	GCC	TGC	CAG	TGC	CCT	GCC	GTG	GTT	TGC	TGC	CGC	TGC	TCA	CCC	ACC	2400
Q	K	A	R	I	V	T	L	L	Q	Q	H	T	G	R	R	T	C	A	I	820
CAG	AAG	GCC	CGC	ATT	GTG	ACA	CTG	CTG	CAG	CAG	CAC	ACA	GGG	AGA	CGC	ACC	TGC	GCC	ATC	2460
G	D	G	G	N	D	V	S	M	I	Q	A	A	D	C	G	I	G	I	E	840
GGT	GAT	GGA	GGA	AAT	GAT	GTC	AGC	ATG	ATT	CAG	GCA	GCA	GAC	TGT	GGG	ATT	GGG	ATT	GAG	2520
G	K	E	G	K	Q	A	S	L	A	A	D	F	S	I	T	Q	F	R	H	860
GGA	AAG	GAG	GGT	AAA	CAG	GCC	TCG	CTG	GCG	GCC	GAC	TTC	TCC	ATC	ACG	CAG	TTC	CGG	CAC	2580
I	G	R	L	L	M	V	H	G	R	N	S	Y	K	R	S	A	A	L	G	880
ATA	GGC	AGG	CTG	CTC	ATG	GTG	CAC	GGG	CGG	AAC	AGC	TAC	AAG	AGG	TCG	GCG	GCA	CTC	GGC	2640
Q	F	V	M	H	R	G	L	I	I	S	T	M	Q	A	V	F	S	S	V	900
CAG	TTC	GTC	ATG	CAC	AGG	GGC	CTT	ATC	ATC	TCC	ACC	ATG	CAG	GCT	GTG	TTT	TCC	TCA	GTC	2700
F	Y	F	A	S	V	P	L	Y	Q	G	F	L	M	V	G	Y	A	T	I	920
TTC	TAC	TTC	GCA	TCC	GTC	CCT	TTG	TAT	CAG	GGC	TTC	CTC	ATG	GTG	GGG	TAT	GCC	ACC	ATA	2760
Y	T	M	F	P	V	F	S	L	V	L	D	Q	D	V	K	P	E	M	A	940
TAC	ACC	ATG	TTC	CCA	GTG	TTC	TCC	TTA	GTG	CTG	GAC	CAG	GAC	GTG	AAG	CCA	GAG	ATG	GCG	2820
M	L	Y	P	E	L	Y	K	D	L	T	K	G	R	S	L	S	F	K	T	960
ATG	CTC	TAC	CCG	GAG	CTG	TAC	AAG	GAC	CTC	ACC	AAG	GGA	AGA	TCC	TTG	TCC	TTC	AAA	ACC	2880
F	L	I	W	V	L	I	S	I	Y	Q	G	G	I	L	M	Y	G	A	L	980
TTC	CTC	ATC	TGG	GTT	TTA	ATA	AGT	ATT	TAC	CAA	GGC	GGC	ATC	CTC	ATG	TAT	GGG	GCC	CTG	2940
V	L	F	E	S	E	F	V	H	V	V	A	I	S	F	T	A	L	I	L	1000
GTG	CTC	TTC	GAG	TCT	GAG	TTC	GTC	CAC	GTG	GTG	GCC	ATC	TCC	TTC	ACC	GCA	CTG	ATC	CTG	3000
T	E	L	L	M	V	A	L	T	V	R	T	W	H	W	L	M	V	V	A	1020
ACC	GAG	CTG	CTG	ATG	GTG	GCG	CTG	ACC	GTC	CGC	ACG	TGG	CAC	TGG	CTG	ATG	GTG	GTG	GCC	3060
E	F	L	S	L	G	C	Y	V	S	S	L	A	F	L	N	E	Y	F	D	1040
GAG	TTC	CTC	AGC	TTA	GGC	TGC	TAC	GTG	TCC	TCA	CTC	GCT	TTT	CTC	AAT	GAA	TAT	TTT	GAT	3120
V	A	F	I	T	T	V	T	F	L	W	K	V	S	A	I	T	V	V	S	1060
GTT	GCC	TTT	ATC	ACC	ACC	GTG	ACC	TTC	CTG	TGG	AAA	GTG	TCG	GCG	ATC	ACC	GTG	GTC	AGC	3180
C	L	P	L	Y	V	L	K	Y	L	R	R	K	S	S	P	P	S	Y	C	1080
TGC	CTC	CCG	CTG	TAT	GTC	CTC	AAG	TAC	CTG	AGG	CGC	AAG	TCT	TCT	CCT	CCC	AGC	TAC	TGC	3240
K	L	A	S	*																1085
AAG	CTG	GCC	TCC	TAA																3255

GGGGCTGTGCACCCCCAGCGGGCTGGCCCCAGCACCTTCTGCCCTTCCCAGCACCTTGTGCCCTTGCCAGTGAACGCAG

FIGURE 32C

GGTTTGCCATTGCTACCAAGCAAGCACCACAAGAAAGGGAGGGTACGCCAGGCGAGCCCAGGGCACAGATGCTGAGACA
GCCTCTCCTTCTCAGTGCAGGGACGTACCCCTGCCAGGCAAGCCCAGGGCACAGATGCCAGGATGGCTTCTCCCTCTC
AGTGCAGGCTTCACCCCTGCCAGGCAAGCCCAGGGCATAGATGCTGAGACAGCCTCTCCCTCTCAGTGCAGGGACGTC
ACCCCTGCCAGGCAAGCCCAGGGCACAGAGGCCGGGACGGCCTCTCCCTCTCAGTGTGAGGCTTCACCCATGCTAGGCA
AGCCCAGGGCACAGATGCCGGGATGGCCCTCCCTCTCAGTGCGGGAACGTACCCCTGCCAGGCAAGCCCAGGGCACA
GATGCTGCGATGGCCTCTTCCTCTTAAGTGTGGGGCCTCACCCCTGCTTTTCTTTCTTTTGTATTGTCAAATTTGT
ATTTCCATATTGAAGCAGCTTGAGTTTCTACTGAAAATGAGCCCGAATTATTTCACTATTACTGTAAAGGGTTCATCTT
ACTCTGGCATTCTGAGAATTAGACTGAAAGTTTAATTTCTGCAGTTCCTCATATTCAGATTCTTTCTTTGATGTTATA
ACACAAAGTCATTCTACTCAAATGTAATAAAATTGAGGCTCCACGGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

FIGURE 32D

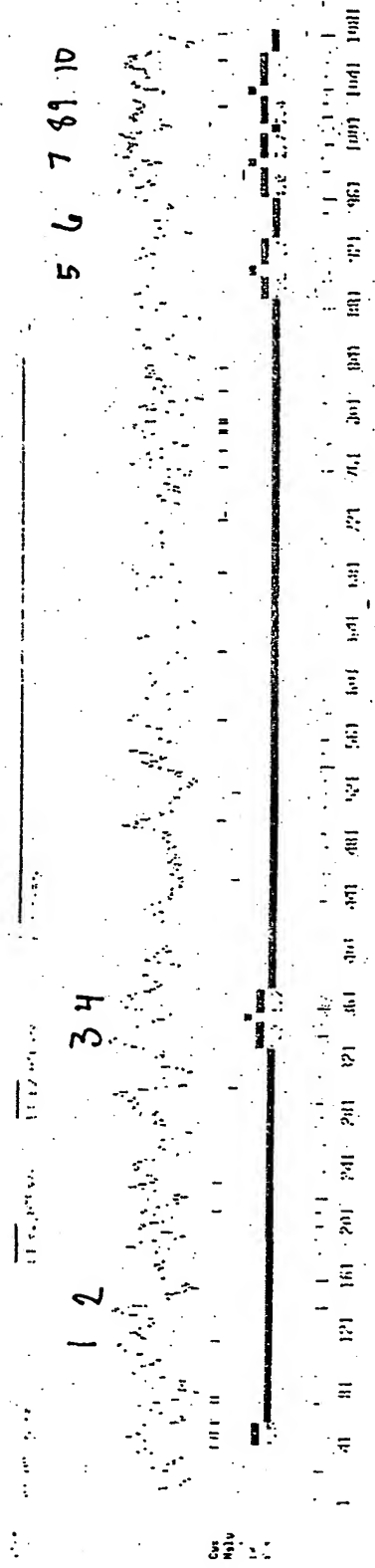


FIGURE 33

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam

Sequence file: /prod/ddm/wspace/orfanal/oa-script.16315.seq

Query: 67084FL

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Hydrolase ¹	haloacid dehalogenase-like hydrolase	19.2	0.0051	
1				
E1-E2_ATPase	E1-E2 ATPase	15.8	0.00087	
2				

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
E1-E2_ATPase	1/2	171	199	42	70	3.0	6.9
E1-E2_ATPase	2/2	277	305	105	133	13.0	0.0064
Hydrolase	1/1	410	843	1	184	19.2	0.0051

Alignments of top-scoring domains:

E1-E2_ATPase: domain 1 of 2, from 171 to 199: score 3.0, E = 6.9

```

*->keeeipaseLvpGDIVevkpGdrVPADgr<-*
+ +++++++GD+++v+ r+P D++
67084FL 171 GKQVQKSSDIQVGDLIIVEKNQRIPSDMV 199

```

E1-E2_ATPase: domain 2 of 2, from 277 to 305: score 13.0, E = 0.0064

```

*->lerngmVfaGTLvvsGsltgvVtatGddT<-*
1 + n++++aT+v sG+ +gvV+ tG++T
67084FL 277 LSIENTLWASTIVASGTVIGVVIYTGKET 305

```

Hydrolase: domain 1 of 1, from 410 to 843: score 19.2, E = 0.0051

```

*->ikavvFDkDGTLTdgkeppiaeaivealrelgl....apleevekl
+ ++ Dk+GTLt+ + i + + +g ++ ++ ++ ++
67084FL 410 LVYLLTDKTGTLTQ--NEMIFKRLHLGTVSYGAdtmdeIQSHVRDSY 454

lgrgl.g.....erilleggltaell.....
+++++++R+++++ + ++ ++ ++ + ++ ++
67084FL 455 SQMQSqAggnntgstpLRKAQSSAPKVRKSVssriheavkaivlchnvtp 504

+++ + +++++ + + + +++++ +++++ +++++ +
67084FL 505 vyesragvteetefaeadqdfdsenrtyqasspdevalvqwtesvgltlv 554

+++ ++ +++++ + +++++ + +++++ +
67084FL 555 srdltsmqlktpsgqvlsfcilqlfpftseskrmgvivrdestaeitfym 604

++ + ++ ++ +++ ++ +++ ++ ++ +++ + +++
67084FL 605 kgadvamspivqyndwleecgnmareglrtlvakkalteeqyqdfesr 654

.....ld.evlglial.dklypgarealkk
+ + + +++++ + ++ +++ e+l+l +d+l ++r++l+ L+
67084FL 655 ytaqlsmhdrslkvaavveslerEmELLCLTGVeDQLQADVVRPTLEMLR 704

```

FIGURE 34A

erGikvailTngdr.naealle.....
+Gik++lT++ ++a+ ++++++ ++++++ + + ++++++ + + +
67084FL 705 NAGIKIWMLTGDKLeTATCIrKsshlvrsrtqdihi frqvtsrgeahleln 754
.....algla.lfdaiivdsdevggvgpvvvgKPkpe
+++++ ++++++ +l + ++++++ +vv+ + +p
67084FL 755 afrrkhdcalvisgdslevCLK-YyEHEFVELACQCP---AVVCCRCSP 800
ifllalerlgvkpeevgpkvlmvGDginDapalaaAGvgvamngg<+
+ +++ l+ + +++++GDg nD+ ++ aA++g+ +
67084FL 801 QKARIVTLLQQHTGRR---TCAIGDGGNDVSMIQWADCGIGIEGKE 843

FIGURE 34B

67084FL
R2B

MPLMMSEEGFENEESDYHTLPRARIMQRKRGLEWFCVCDGWKFCLTSCCGWLINICRRKKE
MPLMMSEEGFENEESDYHTLPRARI TRRKRGLFWCVCWWKFCLTSCDOWLINVCRKKE
*****.*.....*.*****.*.....*:..****

TAN

67084FL
R2B

 LKARTVWLGCPEKCEEKHPRNS IKNQKYNVFTFI PGVLYEQKF FFLNLYFLVISCSQFVP
 LKARTVWLGCPEKCEEKHPRNS IKNQKYNVFTFI PGVLYEQKF FFLNLYFLVVS CSQFVP

7m2

h67084FL
T2B

ALKIGYLYTYWAPLGFVLAVTMTREAIIDEFRFQRDKEVNSQLYSKLTVRGKVQVKSSDI
 ALKIGYLYTYWAPLGFVLAVTIAREAIIDEFRFQRDKEVNSQLYSKLTVRGKVQVKSSDI
 *****:*****:*****

Phospholipid Transport

h67084FL
T2B

QVGDIIIVEKNQRIPSDMVFLRTSEKAGSCFIRIDQLOGETDWWKLKVAVSCTQQLPALGD
QVGDIIIVEKNQRIPSDMVFLRTSEKAGSCFIRIDQLOGETDWWKLKVAVSCTQRLPALGD
*****.*****

h67084FL T2B

LFSISAYVYAQKPQMDIHSFEGTFTRESDPPIHESLSIENTLWASTIVASGTVIGVVIY
LFSISAYVYAQKPQOLDIHSFEGTFTRESDPPIHESLSIENTLWASTIVASGTVIGVVIY

TM3

h67084FL
T2B

TGKETRSVMNTSNPKNKVGLLDLELNRLFKALFLALVALSIVMTLQGFVGWPYRNLFRF
TGKETRSVMNTSNPNNKVGLLDLELNQLFKALFLALVVLSSVMVTLOGFAGWPYRNLFRF

*****.***:*****.*****

TM4

Sh67084FL
AT2B

LLLFSYIIPISLRVNLDMGKAVYGWMMMKDENIPGTVVRTSTIPEELGRLVYLLTOKTGT
 LLLFSYIIPISLRVNLDMGKAAYGWMIMKDENIPGTVVRTSTIPEELGRLVYLLTOKTGT

Phosphoryl
DKTGT 5

5h67084FL
AT2B

L7QNEMIFKRLHLGTVSYGADTMDEIQSHVRDSYSQMQSQAGGNNTGSTPLRKAQSSAPK
L7QNEMVFKRLHLGTVSYGDTMDEIQSHVLNSYLVHSQPSGHNPSSAPLRRSQSSTPK
***** : ** * : ** . * : . : * : **

bh67084FL
AT2B

VRKSVSSRIHEAVKAI VLCHNVTPVYESRAGVTEETFEAEADQDFS DENRTYQASSPDEV
VKKSVSSRIHEAVKAI ALCHNVTPVYEARAGITGETFEAEADQDFS DENRTYQASSPDEV
*****:***:*

bh67084FL
AT2B

ALVQWTESVGLTLVSRDLTSMQLKTPSGQVLSFCILQLFPFTSESKRMGVIVRDESTAEI
ALVRWTESVGLTLVSRDLASMLKTPSGQVLTTCILQMFPPFTSESKRMGIIVRDESTAEI
*****:*****:*****:*****:*****

bh67084FL
AT2B

TFYMKGADVAMSPIVQYNDWLEEECGNMAREGLRTLTVVAKKALTEEQYQDFESRYTOAKL
TFYMKGADVAMSTIVQYNDWLEEECGNMAREGLRTLTVVAKRTLTEEQYQDFESRYSQAKL
***** : ***** : *****

•bh67084FL
1AT2B

SMHDRSLKVAAVVESLEREMELLCLTGVEDQLQADV RPTLEMLRNAGIKIWMLTGDKLET
SIHDRALKVAAVVESLEREMELLCLTGVEDQLQADV RPTLEMLRNAGIKIWMLTGDKLET

bh67084FL
AT2B

ATCIAKSSHLVSRTQDIHI FRQVTSRGEAHLELNAFRRKHDCALVISGDSLEVCLKYEH
ATCIAKSSHLVSRTQDIHVFRPVT SRGEAHLELNAFRRKHDCALVISGDSLEVCLRYYEH

Phospholipid Transport.

Fbh67084FL
MAT2B

EFVELACQCPAVVCCRCSPQKARIVTLLQOHTGRRTCAIGDGGNDVSMIQAADCGIGIE
ELVELACQCPAVVCCRCSPTXKAHIVTLLRQHTKRRTCAIGDGGNDVSMIQAADCGIGIE
***** :***** :***** :***** :***** :***** :*****

TAS

Fbh67084FL

GKEGKOASLAADFSITQFRHIGRLLMVHGRNSYKRSAALGQFVMHRLIISTMQAVFSSV

69/79

AT2B GKEGKQASLAADF SITQFRHIGRLLMVHGRNSYKRSAA LGQFVMHRCGLISTMQAVFSSV

 Tm6
 bh67084FL FYFASVPLYQGF LMGVYATIYTMFPVFSLVLDQDVKPEMAMLYPELYKDLTKGRSLSFKT
 AT2B PYFASVPLYQGF LMGVYATIYTMFPVFSLVLDQDVKPEMAMLYPELYKDLTKGRSLSFKT

 Tm7
 bh67084FL FLIWVLISYQGGILMYGALVLESEFVHVVAISFTALILTELLMVALTVRTWHWLMVVA
 AT2B FLIWVLISYQGGILMYGALVLESEFVHVVAISFTALILTELLXVALTIRTWHLMVVA

 Tm9 Tm10
 bh67084FL EFLSLGCVVSSLAFINEYF-----DVAFITTVTFLWKVSAITVVSCLPLYVLYKY
 AT2B EFLSLGCVVASLAFINEYFGIGRVSGAFLDVAFITTVTFLWKVSAITVVSCLPLYVLYKY

 LRRKSSPPSYCKLAS
 LKRKLSPPSYSKLSS
 :*** *****:*

FIGURE 35B

Input file Fbh67084alt; Output File Fbh67084alt.tra
Sequence length 4231

GGAGTCGACCCACGCGTCCGCATTGAGACAATGCCTCCACAAATACTTGATGCAAAATTCAGTAAGACAGCACTTGTG

AATCACCATTATAGTTTCTGACAAATTGTTCTCAAAAAGGTACCAGCTGGAGGATGAGTCTGCGCATTGATGAA

M	P	L	M	M	S	E	E	G	F	E	N	E	E	S	D	Y	H	T	L	20
ATG	CCA	CTA	ATG	ATG	TCT	GAA	GAA	GGC	TTT	GAG	AAT	GAG	GAA	AGT	GAT	TAC	CAC	ACC	TTA	60
P	R	A	R	I	M	Q	R	K	R	G	L	E	W	F	V	C	D	G	W	40
CCA	CGA	GCC	AGG	ATA	ATG	CAA	AGG	AAA	AGA	GGA	CTG	GAG	TGG	TTT	GTC	TGT	GAT	GGC	TGG	120
K	F	L	C	T	S	C	C	G	W	L	I	N	I	C	R	R	K	K	E	60
AAG	TTC	CTC	TGT	ACC	AGT	TGC	TGT	GGT	TGG	CTG	ATA	AAT	ATT	TGT	CGA	AGA	AAG	AAA	GAG	180
L	K	A	R	T	V	W	L	G	C	P	E	K	C	E	E	K	H	P	R	80
CTG	AAA	GCT	CGC	ACA	GTA	TGG	CTT	GGA	TGT	CCT	GAA	AAG	TGT	GAA	GAA	AAA	CAT	CCC	AGG	240
N	S	I	K	N	Q	K	Y	N	V	F	T	F	I	P	G	V	L	Y	E	100
AAT	TCT	ATA	AAA	AAT	CAA	AAA	TAC	AAT	GTG	TTT	ACC	TTT	ATA	CCT	GGG	GTT	TTG	TAT	GAA	300
Q	F	K	F	F	L	N	L	Y	F	L	V	I	S	C	S	Q	F	V	P	120
CAA	TTC	AAG	TTT	TTT	TTG	AAT	CTC	TAT	TTT	CTA	GTG	ATA	TCC	TGC	TCA	CAG	TTT	GTA	CCA	360
A	L	K	I	G	Y	L	Y	T	Y	W	A	P	L	G	F	V	L	A	V	140
GCA	TTG	AAA	ATA	GGC	TAT	CTC	TAC	ACC	TAC	TGG	GCT	CCT	CTG	GGA	TTT	GTC	TTG	GCT	GTT	420
T	M	T	R	E	A	I	D	E	F	R	R	F	Q	R	D	K	E	V	N	160
ACT	ATG	ACA	CGG	GAA	GCA	ATT	GAT	GAA	TTT	CGG	CGT	TTT	CAG	CGT	GAC	AAG	GAA	GTG	AAT	480
S	Q	L	Y	S	K	L	T	V	R	G	K	V	Q	V	K	S	S	D	I	180
TCA	CAA	CTA	TAT	AGC	AAG	CTT	ACA	GTA	AGA	GGT	AAA	GTG	CAA	GTT	AAG	AGT	TCA	GAC	ATA	540
Q	V	G	D	L	I	I	V	E	K	N	Q	R	I	P	S	D	M	V	F	200
CAA	GTT	GGA	GAC	CTC	ATC	ATA	GTG	GAA	AAG	AAT	CAA	AGA	ATT	CCA	TCG	GAC	ATG	GTG	TTT	600
L	R	T	S	E	K	A	G	S	C	F	I	R	T	D	Q	L	D	G	E	220
CTT	AGG	ACT	TCA	GAA	AAA	GCA	GGT	TCG	TGT	TTT	ATT	CGA	ACT	GAT	CAA	CTA	GAT	GGT	GAA	660
T	D	W	K	L	K	V	A	V	S	C	T	Q	Q	L	P	A	L	G	D	240
ACT	GAC	TGG	AAG	CTG	AAG	GTG	GCA	GTG	AGC	TGC	ACG	CAA	CAG	CTG	CCG	GCT	CTG	GGG	GAC	720
L	F	S	I	S	A	Y	V	Y	A	Q	K	P	Q	M	D	I	H	S	F	260
CTT	TTT	TCT	ATC	AGT	GCT	TAT	GTT	TAT	GCT	CAG	AAA	CCA	CAA	ATG	GAC	ATT	CAC	AGT	TTC	780
E	G	T	F	T	R	E	D	S	D	P	P	I	H	E	S	L	S	I	E	280
GAA	GGC	ACA	TTT	ACC	AGG	GAA	GAC	AGT	GAC	CCG	CCC	ATT	CAT	GAA	AGT	CTC	AGC	ATA	GAA	840
N	T	L	W	A	S	T	I	V	A	S	G	T	V	I	G	V	V	I	Y	300
AAT	ACA	TTG	TGG	GCA	AGC	ACC	ATT	GTT	GCA	TCA	GGT	ACT	GTA	ATA	GGT	GTT	GTC	ATT	TAT	900
T	G	K	E	T	R	S	V	M	N	T	S	N	P	K	N	K	V	G	L	320
ACC	GGA	AAA	GAG	ACT	CGA	AGT	GTA	ATG	AAC	ACA	TCC	AAT	CCA	AAA	AAT	AAG	GTT	GGT	TTG	960
L	D	L	E	L	N	R	L	T	K	A	L	F	L	A	L	V	A	L	S	340
TTG	GAC	CTT	GAA	CTC	AAT	CGG	CTG	ACG	AAA	GCG	CTA	TTT	TTG	GCT	TTA	GTT	GCT	CTT	TCC	1020

FIGURE 36A

I	V	M	V	T	L	Q	G	F	V	G	P	W	Y	R	N	L	F	R	F	360
ATT	GTT	ATG	GTA	ACC	TTA	CAA	GGA	TTT	GTG	GGT	CCA	TGG	TAC	CGC	AAT	CTT	TTT	CGG	TTC	1080
L	L	L	F	S	Y	I	I	P	I	S	L	R	V	N	L	D	M	G	K	380
CTT	CTC	CTC	TTT	TCT	TAC	ATC	ATT	CCC	ATA	AGT	TTG	CGT	GTG	AAC	TTG	GAC	ATG	GGC	AAA	1140
A	V	Y	G	W	M	M	M	K	D	E	N	I	P	G	T	V	V	R	T	400
GCG	GTG	TAT	GGA	TGG	ATG	ATG	ATG	AAA	GAT	GAG	AAC	ATC	CCT	GGC	ACG	GTC	GTT	CGG	ACC	1200
S	T	I	P	E	E	L	G	R	L	V	Y	L	L	T	D	K	T	G	T	420
AGC	ACT	ATC	CCA	GAG	GAA	CTT	GGG	CGC	CTG	GTG	TAT	TTA	TTG	ACA	GAC	AAA	ACA	GGA	ACC	1260
L	T	Q	N	E	M	I	F	K	R	L	H	L	G	T	V	S	Y	G	A	440
CTC	ACC	CAG	AAT	GAA	ATG	ATA	TTT	AAG	CGG	CTG	CAC	CTG	GGC	ACC	GTG	TCC	TAT	GGC	GCC	1320
D	T	M	D	E	I	Q	S	H	V	R	D	S	Y	S	Q	M	Q	S	Q	460
GAC	ACG	ATG	GAT	GAG	ATC	CAG	AGC	CAT	GTC	AGG	GAC	TCC	TAC	TCA	CAG	ATG	CAG	TCT	CAA	1380
A	G	G	N	N	T	G	S	T	P	L	R	K	A	Q	S	S	A	P	K	480
GCT	GGT	GGA	AAC	AAT	ACT	GGT	TCA	ACT	CCA	CTA	AGA	AAA	GCC	CAA	TCT	TCA	GCT	CCC	AAA	1440
V	R	K	S	V	S	S	R	I	H	E	A	V	K	A	I	V	L	C	H	500
GTT	AGG	AAA	AGT	GTC	AGT	AGT	CGA	ATC	CAT	GAA	GCC	GTG	AAA	GCC	ATC	GTG	CTG	TGT	CAC	1500
N	V	T	P	V	Y	E	S	R	A	G	V	T	E	E	T	E	F	A	E	520
AAC	GTG	ACC	CCC	GTG	TAT	GAG	TCT	CGG	GCC	GGC	GTT	ACT	GAG	GAG	ACT	GAG	TTC	GCA	GAG	1560
A	D	Q	D	F	S	D	E	N	R	T	Y	Q	A	S	S	P	D	E	V	540
GCT	GAC	CAA	GAC	TTC	AGT	GAT	GAG	AAT	CGC	ACC	TAC	CAG	GCT	TCC	AGC	CCG	GAT	GAG	GTC	1620
A	L	V	Q	W	T	E	S	V	G	L	T	L	V	S	R	D	L	T	S	560
GCT	CTG	GTG	CAG	TGG	ACA	GAG	AGT	GTG	GGC	CTC	ACG	CTG	GTC	AGC	AGG	GAC	CTC	ACC	TCC	1680
M	Q	L	K	T	P	S	G	Q	V	L	S	F	C	I	L	Q	L	F	P	580
ATG	CAG	CTG	AAG	ACC	CCC	AGT	GGC	CAG	GTC	CTC	AGC	TTC	TGC	ATT	CTG	CAG	CTG	TTT	CCC	1740
F	T	S	E	S	K	R	M	G	V	I	V	R	D	E	S	T	A	E	I	600
TTC	ACC	TCC	GAG	AGC	AAG	CGG	ATG	GGC	GTC	ATC	GTC	AGG	GAT	GAA	TCC	ACG	GCA	GAA	ATC	1800
T	F	Y	M	K	G	A	D	V	A	M	S	P	I	V	Q	Y	N	D	W	620
ACA	TTC	TAC	ATG	AAG	GGC	GCT	GAC	GTG	GCC	ATG	TCT	CCT	ATC	GTG	CAG	TAT	AAT	GAC	TGG	1860
L	E	E	E	C	G	N	M	A	R	E	G	L	R	T	L	V	V	A	K	640
CTG	GAA	GAG	GAG	TGC	GGA	AAC	ATG	GCT	CGC	GAA	GGA	CTG	CGG	ACC	CTC	GTG	GTT	GCA	AAG	1920
K	A	L	T	E	E	Q	Y	Q	D	F	E	S	R	Y	T	Q	A	K	L	660
AAG	GCG	TTG	ACA	GAG	GAG	CAG	TAC	CAG	GAC	TTT	GAG	AGC	CGA	TAC	ACT	CAA	GCC	AAG	CTG	1980
S	M	H	D	R	S	L	K	V	A	A	V	V	E	S	L	E	R	E	M	680
AGC	ATG	CAC	GAC	AGG	TCC	CTC	AAG	GTG	GCC	GCG	GTA	GTC	GAG	AGC	CTG	GAG	AGG	GAG	ATG	2040
E	L	L	C	L	T	G	V	E	D	Q	L	Q	A	D	V	R	P	T	L	700
GAA	CTG	CTG	TGC	CTC	ACC	GGC	GTG	GAG	GAC	CAG	CTG	CAG	GCA	GAC	GTG	CGG	CCC	ACG	CTG	2100
E	M	L	R	N	A	G	I	K	I	W	M	L	T	G	D	K	L	E	T	720
GAG	ATG	CTG	CGC	AAC	GCC	GGG	ATC	AAG	ATA	TGG	ATG	CTA	ACA	GGC	GAT	AAA	CTC	GAG	ACA	2160

FIGURE 36B

A	T	C	I	A	K	S	S	H	L	V	S	R	T	Q	D	I	H	I	F	740
GCT	ACC	TGC	ATT	GCC	AAA	AGT	TCA	CAT	CTC	GTG	TCT	AGA	ACA	CAA	GAT	ATT	CAT	ATT	TTC	2220
R	Q	V	T	S	R	G	E	A	H	L	E	L	N	A	F	R	R	K	H	760
AGA	CAG	GTA	ACC	AGT	CGG	GGA	GAG	GCA	CAT	TTG	GAG	CTG	AAT	GCA	TTT	CGA	AGG	AAG	CAT	2280
D	C	A	L	V	I	S	G	D	S	L	E	V	C	L	K	Y	Y	E	H	780
GAT	TGT	GCA	CTA	GTC	ATA	TCT	GGG	GAC	TCT	CTG	GAG	GTT	TGT	CTA	AAG	TAC	TAC	GAG	CAT	2340
E	F	V	E	L	A	C	Q	C	P	A	V	V	C	C	R	C	S	F	T	800
GAA	TTT	GTG	GAG	CTG	GCC	TGC	CAG	TGC	CCT	GCC	GTG	GTT	TGC	TGC	CGC	TGC	TCA	CCC	ACC	2400
Q	K	A	R	I	V	T	L	L	Q	Q	H	T	G	R	R	T	C	A	I	820
CAG	AAG	GCC	CGC	ATT	GTG	ACA	CTG	CTG	CAG	CAG	CAC	ACA	GGG	AGA	CGC	ACC	TGC	GCC	ATC	2460
G	D	G	G	N	D	V	S	M	I	Q	A	A	D	C	G	I	G	I	E	840
GGT	GAT	GGA	GGA	AAT	GAT	GTC	AGC	ATG	ATT	CAG	GCA	GCA	GAC	TGT	GGG	ATT	GGG	ATT	GAG	2520
G	K	E	G	K	Q	A	S	L	A	A	D	F	S	I	T	Q	F	R	H	860
GGA	AAG	GAG	GGT	AAA	CAG	GCC	TCG	CTG	GCG	GCC	GAC	TTC	TCC	ATC	ACG	CAG	TTC	CGG	CAC	2580
I	G	R	L	L	M	V	H	G	R	N	S	Y	K	R	S	A	A	L	G	880
ATA	GGC	AGG	CTG	CTC	ATG	GTG	CAC	GGG	CGG	AAC	AGC	TAC	AAG	AGG	TCG	GCG	GCA	CTC	GGC	2640
Q	F	V	M	H	R	G	L	I	I	S	T	M	Q	A	V	F	S	S	V	900
CAG	TTC	GTC	ATG	CAC	AGG	GGC	CTT	ATC	ATC	TCC	ACC	ATG	CAG	GCT	GTG	TTT	TCC	TCA	GTC	2700
F	Y	F	A	S	V	P	L	Y	Q	G	F	L	M	V	G	Y	A	T	I	920
TTC	TAC	TTC	GCA	TCC	GTC	CCT	TTG	TAT	CAG	GGC	TTC	CTC	ATG	GTG	GGG	TAT	GCC	ACC	ATA	2760
Y	T	M	F	P	V	F	S	L	V	L	D	Q	D	V	K	P	E	M	A	940
TAC	ACC	ATG	TTC	CCA	GTG	TTC	TCC	TTA	GTG	CTG	GAC	CAG	GAC	GTG	AAG	CCA	GAG	ATG	GCG	2820
M	L	Y	P	E	L	Y	K	D	L	T	K	G	R	S	L	S	F	K	T	960
ATG	CTC	TAC	CCG	GAG	CTG	TAC	AAG	GAC	CTC	ACC	AAG	GGA	AGA	TCC	TTG	TCC	TTC	AAA	ACC	2880
F	L	I	W	V	L	I	S	I	Y	Q	G	G	I	L	M	Y	G	A	L	980
TTC	CTC	ATC	TGG	GTT	TTA	ATA	AGT	ATT	TAC	CAA	GGC	GGC	ATC	CTC	ATG	TAT	GGG	GCC	CTG	2940
V	L	F	E	S	E	F	V	H	V	V	A	I	S	F	T	A	L	I	L	1000
GTG	CTC	TTC	GAG	TCT	GAG	TTC	GTC	CAC	GTG	GTG	GCC	ATC	TCC	TTC	ACC	GCA	CTG	ATC	CTG	3000
T	E	L	L	M	V	A	L	T	V	R	T	W	H	W	L	M	V	V	A	1020
ACC	GAG	CTG	CTG	ATG	GTG	GCG	CTG	ACC	GTC	CGC	ACG	TGG	CAC	TGG	CTG	ATG	GTG	GTG	GCC	3060
E	F	L	S	L	G	C	Y	V	S	S	L	A	F	L	N	E	Y	F	G	1040
GAG	TTC	CTC	AGC	TTA	GGC	TGC	TAC	GTG	TCC	TCA	CTC	GCT	TTT	CTC	AAT	GAA	TAT	TTT	GGT	3120
I	G	R	V	S	F	G	A	F	L	D	V	A	F	I	T	T	V	T	F	1060
ATA	GGC	AGA	GTG	TCT	TTT	GGA	GCT	TTC	TTA	GAT	GTT	GCC	TTT	ATC	ACC	ACC	GTG	ACC	TTC	3180
L	W	K	V	S	A	I	T	V	V	S	C	L	P	L	Y	V	L	K	Y	1080
CTG	TGG	AAA	GTG	TCG	GCG	ATC	ACC	GTG	GTC	AGC	TGC	CTC	CCG	CTG	TAT	GTC	CTC	AAG	TAC	3240
L	R	R	K	S	S	P	P	S	Y	C	K	L	A	S	*					1096
CTG	AGG	CGC	AAG	TCT	TCT	CCT	CCC	AGC	TAC	TGC	AAG	CTG	GCC	TCC	TAA					3288

GGGGCTGTGCACCCCCAGCGGGCTGGCCCCAGCACCTTCTGCCCTTCCCAGCACCTTGTGCCCTTGCCAGTGAACGCAG

FIGURE 36C

GGTTTGCCATTGCTACCAAGCAAGCACCACAAGAAAGGGAGGGTACGCCAGGCGAGCCCAGGGCACAGATGCTGAGACA
GCCTCTCCTTCTCAGTGCAGGGACGTACCCCTGCCAGGCAAGCCCAGGGCACAGATGCCAGGATGGCTTCTCCCTCTC
AGTGCGAGGGCTTCACCCCTGCCAGGCAAGCCCAGGGCATAGATGCTGAGACAGCTCTCCCTCTCAGTGCAGGGACGTC
ACCCCTGCCAGGCAAGCCCAGGGCACAGAGGCCGGGACGGCCTCTCCCTCTCAGTGTGAGGCTTCACCCATGCTAGGCA
AGCCCAGGGCACAGATGCCGGGATGGCCTCTCCCTCTCAGTGCAGGGAACGTACCCCTGCCAGGCAAGCCCAGGGCAC
GATGCTGCGATGGCCTCTTCTCTTAAGTGTGGGGCCTCACCCCTGCTTTTCTTTTGTATTGTCAAATGT
ATTTCCATATTGAAGCAGCTTGAGTTTCTACTGAAAATGAGCCGAATTATTTCACTATTACTGTAAAGGGTTCATCTT
ACTCTGGCATTCTGAGAAATAGACTGAAAGTTTAATTTCTGCAGTTCCTCATATTCAGATTCTTTCTTTGATGTTATA
ACACAAAGTCATTCTACTCAAATGTAATAAAATTGAGGCTCCACGGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAA

FIGURE 36D

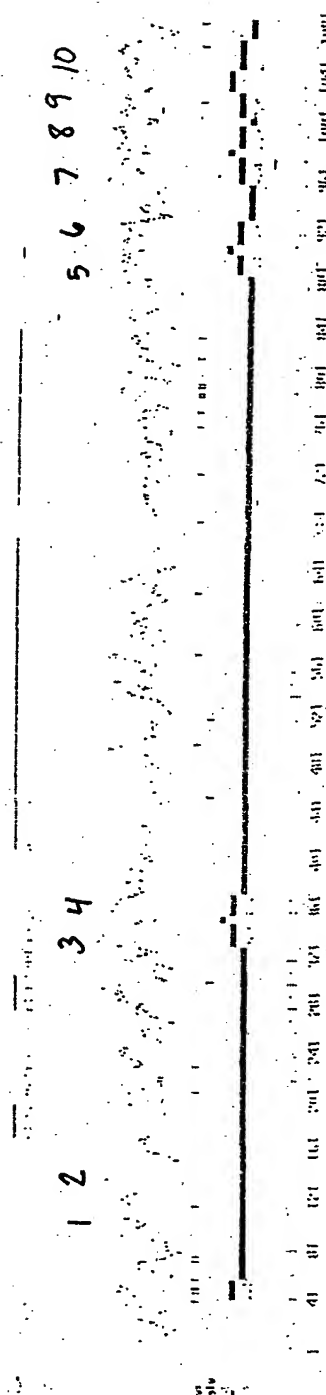


FIGURE 37

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam

Sequence file: /prod/ddm/wspace/orfanal/oa-script.17118.seq

Query: 67084alt

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Hydrolase	haloacid dehalogenase-like hydrolase	19.2	0.0051	
1				
E1-E2_ATPase	E1-E2 ATPase	15.8	0.00087	
2				

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
E1-E2_ATPase	1/2	171	199	42	70	3.0	6.9
E1-E2_ATPase	2/2	277	305	105	133	13.0	0.0064
Hydrolase	1/1	410	843	1	184	19.2	0.0051

Alignments of top-scoring domains:

E1-E2_ATPase: domain 1 of 2, from 171 to 199: score 3.0, E = 6.9

```

*->keeeipaeelvpGDIVevkpGdrVPADgr<-*
+ +++++++GD+++v+ r+P D++
67084alt 171 GKVVQKSSDIQVGDIIIVEKNQRIPSDMV 199

```

E1-E2_ATPase: domain 2 of 2, from 277 to 305: score 13.0, E = 0.0064

```

*->lergnmVfaGTLlvvsGsltgVtGddT<-*
1 + n+++a+T+v sG+ +gvV+ tG++T
67084alt 277 LSIENLWASTIVASGTVIGVVIYTGKET 305

```

Hydrolase: domain 1 of 1, from 410 to 843: score 19.2, E = 0.0051

```

*->ikavvFDkDGTltdgkeppiaaeivealrelgl.....apleevekl
+ ++ Dk+GTLt+ + i + + +g ++ ++ ++ ++
67084alt 410 LVYLLTDKGTGLTQ--NEMIFKRLHLGTVSYGAdtmdeIQSHVRDSY 454

lgrgl.g.....erilleggltaell.....
+++++++I+++++ + ++ ++ ++ + ++ ++
67084alt 455 SQMQSgAggnntgstpLRKAQSSAPKVRKsvssriheavkaivlchnvtp 504

+++ + +++++ + + + +++++ +++++ + + +
67084alt 505 vyesragvteetefaeadqdfsdenrtyqasspdevalvqwtesvgltlv 554

+++ ++ +++++ + + +++++ + +++++ + +
67084alt 555 srdltsmqlktpsgqvlscilqlfpftseskrmgvivrdestaetfym 604

++ + ++ ++ ++ ++ ++ ++ ++ ++ ++ ++
67084alt 605 kgadvamspivqyndwleeecgnmareglrtlvakkalteeqyqdfesr 654

.....ld.evlglial.dklypgarealkalk
+ + + +++++ + ++ ++ e+l+l ++d+l ++r++l+ L+
67084alt 655 ytgaklsmhdrslkvaavvesleREmELLCLTGVeDQLQADVRPTLEMLR 704

```

FIGURE 38A

```
erGikvailTngdr.naealle.....
+Gik++lT++ ++a+ +++++ +++++ + ++ +++++ + + +
67084alt 705 NAGIKIWMLTGDKLeTATCIAKsshlvstqdihi frqvt srgeahleln 754
.....algla.lfdai vdsdevggvgpvvvgKPkpe
+++++ +++++ + +l + +++++ +++++ +vv+ + +p
67084alt 755 afrkhdcalvisgdslevCLK-YyEHEFVELACQCP---AVVCCRCSPt 800
ifllalerlgvkpeevgpkvlmvGDginDapalaaAGvgvamngg<-*
+ + + + l+ + +++++GDg nD+ ++ aA++g+ +
67084alt 801 QKARIVTLLQOHTGRR---TCAIGDGGNDVSMIQAADCGIGIEGKE 843
```

FIGURE 38B

CLUSTAL W (1.74) multiple sequence alignment

```

Fbh67084alt  MPLMMSEEGFENEESDYHTLPRARIMQRKRGLEWFVCDGWKFLCTSCCGWLINICRRKKE
mAT2B          MPLMMSEEGFENDES DYHTLPRARITRRKRGLEWFVCGGWKFLCTSCCDWLINVCQRKKE
                *****:*****:*****:*****:*****:*****:*****:*****
                TM1
Fbh67084alt  LKARTVWLGCPEKCEEKHPRNSIKNQKYNVFTFIPGVLYEOKFELNLYFLVISCQSOFVP
mAT2B          LKARTVWLGCPEKCEEKHPRNSIKNQKYNVFTFIPGVLYEOKFELNLYFLVVSQSOFVP
                *****:*****:*****:*****:*****:*****:*****
                TM2
Fbh67084alt  ALKIGYLYTYWAPLGFVLAVTMTREAIDEFRRFQDKVNSQLYSKLTVRGKVQVKSSDI
mAT2B          ALKIGYLYTYWAPLGFVLAVTIAREAIDEFRRFQDKEMNSQLYSKLTVRGKVQVKSSDI
                *****:*****:*****:*****:*****:*****:*****
                phospholipid transport
Fbh67084alt  QVGDLIIVEKNQRIPSDMVFLRTSEKAGSCFIRIDQIDGETDWKLKVAVSCTQQLPALGD
mAT2B          QVGDLIIVEKNQRIPSDMVFLRTSEKAGSCFIRIDQIDGETDWKLKVAVSCTQRLPALGD
                *****:*****:*****:*****:*****:*****:*****
Fbh67084alt  LFSISAYVYAQKPQMDIHSFEGFTTREDSDPPIHESLSIENTLWASTIVASGTVIGVVIY
mAT2B          LFSISAYVYAQKPLDIHSFEGFTTREDSDPPIHESLSIENTLWASTIVASGTVIGVVIY
                *****:*****:*****:*****:*****:*****:*****
                TM3
Fbh67084alt  TGKETRSVMNTSNPKNKVGLLDLELNRLTKALFLALVALSIVMVTLOGFVGPWYRNLEFRF
mAT2B          TGKETRSVMNTSNPNNKVGLLDLELNQITKALFLALVLSVVMVTLOGFAGPWYRNLEFRF
                *****:*****:*****:*****:*****:*****:*****
                TM4
Fbh67084alt  LLLFSYIIPISLRVNLDMGKAVYGWMMMKDENIPGTVVRTSTIPEELGRLVYLLTDKTGT
mAT2B          LLLFSYIIPISLRVNLDMOKAAYGWMIMKDENIPGTVVRTSTIPEELGRLVYLLTDKTGT
                *****:*****:*****:*****:*****:*****:*****
                phosphoryl. st
Fbh67084alt  LTQNEMIFKRLHLGTVSYGADTMDEIQSHVRDSYSQMOSQAGGNNTGSTPLRKAQSSAPK
mAT2B          LTQNEMVFKRLHLGTVSYGDTMDEIQSHVLNSYLQVHSQPSGHNPPSAPLRRSQSSTPK
                *****:*****:*****:*****:*****:*****:*****
Fbh67084alt  VRKSVSSRIHEAVKAIVLCHNVTPVYESRAGVTEETFAEADQDFSDENRTYQASSPDEV
mAT2B          VKKSVSSRIHEAVKAIALCHNVTPVYEARAGITGETFAEADQDFSDENRTYQASSPDEV
                *:*****:*****:*****:*****:*****:*****:*****
Fbh67084alt  ALVQWTESVGLTLVSRDLTSMQLKTPSGQVLSFCILQLFPFTSESKRMGVIVRDESTAEI
mAT2B          ALVRWTESVGLTLVSRDLASMQLKTPSGQVLTICYILQMFPTSESKRMGIIVRDESTAEI
                *:*****:*****:*****:*****:*****:*****:*****
Fbh67084alt  TFYMKGADVAMSPIVQYNDWLEEECGNMAREGLRTLVAKKALTEEQYQDFESRYTOAKL
mAT2B          TFYMKGADVAMSTIVQYNDWLEEECGNMAREGLRTLVAKRTLTEEQYQDFESRYSOAKL
                *****:*****:*****:*****:*****:*****:*****
Fbh67084alt  SMHDRSLKVAAVVESLEREMELLCLTGVEDQLQADV RPTLEMLRNAGIKIWMLTGDKLET
mAT2B          SIHDRALKVAAVVESLEREMELLCLTGVEDQLQADV RPTLEMLRNAGIKIWMLTGDKLET
                *:*****:*****:*****:*****:*****:*****:*****
Fbh67084alt  ATCIAKSSHLVSRTQDIHIFRQVTSRGEAHLELNAFRKHDCAVISGDSLEVCLKYEH
mAT2B          ATCIAKSSHLVSRTQDIHVFRPVTSRGEAHLELNAFRKHDCAVISGDSLEVCLRYEH
                *****:*****:*****:*****:*****:*****:*****
                phospholipid transport
Fbh67084alt  EFVELACQCPAVVCCRCSPQKARIVTLLQOHTGRRTCAIGDGGNDVSMIQAADCGIGIE
mAT2B          ELVELACQCPAVVCCRCSPYKAHIVTLLRQHTKRTCAIGDGGNDVSMIQAADCGIGIE
                *:*****:*****:*****:*****:*****:*****:*****
                TM5
Fbh67084alt  GKEGKQASLAADFSITQFRHIGRLLMVHGRNSYKRS AALGQFVMHRLIISTMQAVFSSV

```

FIGURE 39A

AT2B GKEGKQASLAADFSITQFRHIGRLLMVHGRNSYKRSAA LGQFVMHRGLIISTMQAVFSSV

TM 6

bh67084alt
AT2B FYFASVPLYQCFFLMVGYATIYTMFPVFSLVLDQDVKPEMAMLYPELYKDLTKGRSLSFKT
FYFASVPLYQCFFLMVGYATIYTMFPVFSLVLDQDVKPEMAMLYPELYKDLTKGRSLSFKT

bh67084alt
AT2B FLIWVLISIIYQGGILMYGALVLEEFSEFVHVVAISFTALILTELLMVALTVRTWHLMVVA
FLIWVLISIIYQGGILMYGAILLFEDEFVHVVAISFTALILTELLXVALTIRTWHLMVVA

TM 7 TM 8

bh67084alt
AT2B EFLSLGCVSSLAFLNEYFGIGRVSFGAFLDVAFITTVTFLWKVSAITVVSCLPYVVKY
EFLSLGCVASLAFLNEYFGIGRVSFGAFLDVAFITTVTFLWKVSAITVVSCLPYVVKY

TM 9 TM 10

bh67084alt
AT2B LRRKSSPPSYCKLAS
LKRKLSPPSYSKLSS
::* *:*:*

FIGURE 39B

SEQUENCE LISTING

<110> Millennium Pharmaceuticals, Inc.

~~Curtis, Rory A.J.~~

<120> 8099, 146455, 54414, 53763, 67076, 67102, 44181,
67084FL, AND 67084 ALT, HUMAN PROTEINS AND METHODS OF
USE THEREOF

<130> MNI-214CPPC

<150> US 60/256,240

<151> 2000-12-15

<150> US 60/256,588

<151> 2000-12-18

<150> US 60/258,028

<151> 2000-12-21

<160> 40

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2725

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (180)...(2033)

<400> 1

```

ccacgcgtcc ggccttccga aatagaaaca aagttggtca caaatcacat tagctttgcc 60
cgaagttttt cccacactc ttcttttagca tgctattatg gggaaagtga ccactcctgg 120
gagcgggggt ggtcggggcg gtttggtggc ggggaagcgg ctgtaacttc tacgtgacc 179
atg gta cct gtt gaa aac acc gag ggc ccc agt ctg ctg aac cag aag 227
Met Val Pro Val Glu Asn Thr Glu Gly Pro Ser Leu Leu Asn Gln Lys
  1           5           10           15

ggg aca gcc gtg gag acg gag ggc agc ggc agc cgg cat cct ccc tgg 275
Gly Thr Ala Val Glu Thr Glu Gly Ser Gly Ser Arg His Pro Pro Trp
          20           25           30

gcg aga ggc tgc ggc atg ttt acc ttc ctg tca tct gtc act gct gct 323
Ala Arg Gly Cys Gly Met Phe Thr Phe Leu Ser Ser Val Thr Ala Ala
          35           40           45

gtc agt ggc ctc ctg gtg ggt tat gaa ctt ggg atc atc tct ggg gct 371
Val Ser Gly Leu Leu Val Gly Tyr Glu Leu Gly Ile Ile Ser Gly Ala
          50           55           60

ctt ctt cag atc aaa acc tta tta gcc ctg agc tgc cat gag cag gaa 419
Leu Leu Gln Ile Lys Thr Leu Leu Ala Leu Ser Cys His Glu Gln Glu
          65           70           75           80

atg gtt gtg agc tcc ctc gtc att gga gcc ctc ctt gcc tca ctc acc 467
Met Val Val Ser Ser Leu Val Ile Gly Ala Leu Leu Ala Ser Leu Thr
          85           90           95

```

gga ggg gtc ctg ata gac aga tat gga aga agg aca gca atc atc ttg	515
Gly Gly Val Leu Ile Asp Arg Tyr Gly Arg Arg Thr Ala Ile Ile Leu	
100 105 110	
tca tcc tgc ctg ctt gga ctc gga agc tta gtc ttg atc ctc agt tta	563
Ser Ser Cys Leu Leu Gly Leu Gly Ser Leu Val Leu Ile Leu Ser Leu	
115 120 125	
tcc tac acg gtt ctt ata gtg gga cgc att gcc ata ggg gtc tcc atc	611
Ser Tyr Thr Val Leu Ile Val Gly Arg Ile Ala Ile Gly Val Ser Ile	
130 135 140	
tcc ctc tct tcc att gcc act tgt gtt tac atc gca gag att gct cct	659
Ser Leu Ser Ser Ile Ala Thr Cys Val Tyr Ile Ala Glu Ile Ala Pro	
145 150 155 160	
caa cac aga aga ggc ctt ctt gtg tca ctg aat gag ctg atg att gtc	707
Gln His Arg Arg Gly Leu Leu Val Ser Leu Asn Glu Leu Met Ile Val	
165 170 175	
atc ggc att ctt tct gcc tat att tca aat tac gca ttt gct aat gtt	755
Ile Gly Ile Leu Ser Ala Tyr Ile Ser Ash Tyr Ala Phe Ala Asn Val	
180 185 190	
ttc cat ggc tgg aag tac atg ttt ggt ctt gtg att ccc ttg gga gtt	803
Phe His Gly Trp Lys Tyr Met Phe Gly Leu Val Ile Pro Leu Gly Val	
195 200 205	
ttg caa gca att gca atg tat ttt ctt cct cca agc cct cgg ttt ctg	851
Leu Gln Ala Ile Ala Met Tyr Phe Leu Pro Pro Ser Pro Arg Phe Leu	
210 215 220	
gtg atg aaa gga caa gag gga gct gct agc aag gtt ctt gga agg tta	899
Val Met Lys Gly Gln Glu Gly Ala Ala Ser Lys Val Leu Gly Arg Leu	
225 230 235 240	
aga gca ctc tca gat aca act gag gaa ctc act gtg atc aaa tcc tcc	947
Arg Ala Leu Ser Asp Thr Thr Glu Glu Leu Thr Val Ile Lys Ser Ser	
245 250 255	
ctg aaa gat gaa tat cag tac agt ttt tgg gat ctg ttt cgt tca aaa	995
Leu Lys Asp Glu Tyr Gln Tyr Ser Phe Trp Asp Leu Phe Arg Ser Lys	
260 265 270	
gac aac atg cgg acc cga ata atg ata gga cta aca cta gta ttt ttt	1043
Asp Asn Met Arg Thr Arg Ile Met Ile Gly Leu Thr Leu Val Phe Phe	
275 280 285	
gta caa atc act ggc caa cca aac ata ttg ttc tat gca tca act gtt	1091
Val Gln Ile Thr Gly Gln Pro Asn Ile Leu Phe Tyr Ala Ser Thr Val	
290 295 300	
ttg aag tca gtt gga ttt caa agc aat gag gca gct agc ctc gcc tcc	1139
Leu Lys Ser Val Gly Phe Gln Ser Asn Glu Ala Ala Ser Leu Ala Ser	
305 310 315 320	
act ggg gtt gga gtc gtc aag gtc att agc acc atc cct gcc act ctt	1187
Thr Gly Val Gly Val Val Lys Val Ile Ser Thr Ile Pro Ala Thr Leu	
325 330 335	

ctt gta gac cat gtc ggc agc aaa aca ttc ctc tgc att ggc tcc tct	1235
Leu Val Asp His Val Gly Ser Lys Thr Phe Leu Cys Ile Gly Ser Ser	
340 345 350	
gtg atg gca gct tcg ttg gtg acc atg ggc atc gta aat ctc aac atc	1283
Val Met Ala Ala Ser Leu Val Thr Met Gly Ile Val Asn Leu Asn Ile	
355 360 365	
cac atg aac ttc acc cat atc tgc aga agc cac aat tct atc aac cag	1331
His Met Asn Phe Thr His Ile Cys Arg Ser His Asn Ser Ile Asn Gln	
370 375 380	
tcc ttg gat gag tct gtg att tat gga cca gga aac ctg tca acc aac	1379
Ser Leu Asp Glu Ser Val Ile Tyr Gly Pro Gly Asn Leu Ser Thr Asn	
385 390 395 400	
aac aat act ctc aga gac cac ttc aaa ggg att tct tcc cat agc aga	1427
Asn Asn Thr Leu Arg Asp His Phe Lys Gly Ile Ser Ser His Ser Arg	
405 410 415	
agc tca ctc atg ccc ctg aga aat gat gtg gat aag aga ggg gag acg	1475
Ser Ser Leu Met Pro Leu Arg Asn Asp Val Asp Lys Arg Gly Glu Thr	
420 425 430	
acc tca gca tcc ttg cta aat gct gga tta agc cac act gaa tac cag	1523
Thr Ser Ala Ser Leu Leu Asn Ala Gly Leu Ser His Thr Glu Tyr Gln	
435 440 445	
ata gtc aca gac cct ggg gac gtc cca gct ttt ttg aaa tgg ctg tcc	1571
Ile Val Thr Asp Pro Gly Asp Val Pro Ala Phe Leu Lys Trp Leu Ser	
450 455 460	
tta gcc agc ttg ctt gtt tat gtt gct gct ttt tca att ggt cta gga	1619
Leu Ala Ser Leu Leu Val Tyr Val Ala Ala Phe Ser Ile Gly Leu Gly	
465 470 475 480	
cca atg ccc tgg ctg gtg ctc agc gag atc ttt cct ggt ggg atc aga	1667
Pro Met Pro Trp Leu Val Leu Ser Glu Ile Phe Pro Gly Gly Ile Arg	
485 490 495	
gga cga gcc atg gct tta act tct agc atg aac tgg ggc atc aat ctc	1715
Gly Arg Ala Met Ala Leu Thr Ser Ser Met Asn Trp Gly Ile Asn Leu	
500 505 510	
ctc atc tcg ctg aca ttt ttg act gta act gat ctt att ggc ctg cca	1763
Leu Ile Ser Leu Thr Phe Leu Thr Val Thr Asp Leu Ile Gly Leu Pro	
515 520 525	
tgg gtg tgc ttt ata tat aca atc atg agt cta gca tcc ctg ctt ttt	1811
Trp Val Cys Phe Ile Tyr Thr Ile Met Ser Leu Ala Ser Leu Leu Phe	
530 535 540	
gtt gtt atg ttt ata cct gag aca aag gga tgc tct ttg gaa caa ata	1859
Val Val Met Phe Ile Pro Glu Thr Lys Gly Cys Ser Leu Glu Gln Ile	
545 550 555 560	
tca atg gag cta gca aaa gtg aac tat gtg aaa aac aac att tgt ttt	1907
Ser Met Glu Leu Ala Lys Val Asn Tyr Val Lys Asn Asn Ile Cys Phe	
565 570 575	

atg agt cat cac caa gaa gaa tta gtg cca aaa cag cct caa aaa aga 1955
 Met Ser His His Gln Glu Glu Leu Val Pro Lys Gln Pro Gln Lys Arg
 580 585 590

aaa ccc cag gag cag ctc ttg gag tgt aac aag ctg tgt ggt agg ggc 2003
 Lys Pro Gln Glu Gln Leu Leu Glu Cys Asn Lys Leu Cys Gly Arg Gly
 595 600 605

caa tcc agg cag ctt tct cca gag acc taa tggcctcaac accttctgaa 2053
 Gln Ser Arg Gln Leu Ser Pro Glu Thr *
 610 615

cggtggatagt gccagaacac ttagggagggt gtctttggac caatgcatag ttgcgactcc 2113
 tgtgtctctct ttccagtgtc atggaactgg ttttgaagag acactctgaa atgataaaga 2173
 cagccttttaa tccccctcct cccagaagg aacctcaaaa ggtagatgag gtacaagggtc 2233
 ctaagtgatc tctttttctg agcaggatat cagggttaaaa aaaaaaagtt actggctgggt 2293
 ttaatacttt ctaccttctt cacagagcag cctttgaata gactatgtcc tagtgaagac 2353
 atcaacctcc gccttaagct atgtatgtat ggaggccagt cgcagcttta ttatgcagac 2413
 acacaagtgg tctggacatg agggtagact ttctgcctac caagacacta cttgcactgg 2473
 atcttacgca aaaaagaacc agaacacaca gtgtggacaa ctgcccataat attctatcta 2533
 gattaggaga gggtcctggc taggatttta gtggtaattc ctagttacat tcaacaagta 2593
 taaagattat agagcttatt ttatgaacta taaactataa tttaatgcaa aatatccttt 2653
 tatgaatttc atgttaatat tgtgaaatat taaaataatt ccgcaataaa aaaaaaaaaa 2713
 aagggcggcc gc 2725

<210> 2

<211> 617

<212> PRT

<213> Homo sapiens

<400> 2

Met Val Pro Val Glu Asn Thr Glu Gly Pro Ser Leu Leu Asn Gln Lys
 1 5 10 15
 Gly Thr Ala Val Glu Thr Glu Gly Ser Gly Ser Arg His Pro Pro Trp
 20 25 30
 Ala Arg Gly Cys Gly Met Phe Thr Phe Leu Ser Ser Val Thr Ala Ala
 35 40 45
 Val Ser Gly Leu Leu Val Gly Tyr Glu Leu Gly Ile Ile Ser Gly Ala
 50 55 60
 Leu Leu Gln Ile Lys Thr Leu Leu Ala Leu Ser Cys His Glu Gln Glu
 65 70 75 80
 Met Val Val Ser Ser Leu Val Ile Gly Ala Leu Leu Ala Ser Leu Thr
 85 90 95
 Gly Gly Val Leu Ile Asp Arg Tyr Gly Arg Arg Thr Ala Ile Ile Leu
 100 105 110
 Ser Ser Cys Leu Leu Gly Leu Gly Ser Leu Val Leu Ile Leu Ser Leu
 115 120 125
 Ser Tyr Thr Val Leu Ile Val Gly Arg Ile Ala Ile Gly Val Ser Ile
 130 135 140
 Ser Leu Ser Ser Ile Ala Thr Cys Val Tyr Ile Ala Glu Ile Ala Pro
 145 150 155 160
 Gln His Arg Arg Gly Leu Leu Val Ser Leu Asn Glu Leu Met Ile Val
 165 170 175
 Ile Gly Ile Leu Ser Ala Tyr Ile Ser Asn Tyr Ala Phe Ala Asn Val
 180 185 190
 Phe His Gly Trp Lys Tyr Met Phe Gly Leu Val Ile Pro Leu Gly Val
 195 200 205
 Leu Gln Ala Ile Ala Met Tyr Phe Leu Pro Pro Ser Pro Arg Phe Leu
 210 215 220
 Val Met Lys Gly Gln Glu Gly Ala Ala Ser Lys Val Leu Gly Arg Leu
 225 230 235 240

Arg	Ala	Leu	Ser	Asp	Thr	Thr	Glu	Glu	Leu	Thr	Val	Ile	Lys	Ser	Ser
				245					250					255	
Leu	Lys	Asp	Glu	Tyr	Gln	Tyr	Ser	Phe	Trp	Asp	Leu	Phe	Arg	Ser	Lys
			260					265					270		
Asp	Asn	Met	Arg	Thr	Arg	Ile	Met	Ile	Gly	Leu	Thr	Leu	Val	Phe	Phe
		275					280					285			
Val	Gln	Ile	Thr	Gly	Gln	Pro	Asn	Ile	Leu	Phe	Tyr	Ala	Ser	Thr	Val
	290					295					300				
Leu	Lys	Ser	Val	Gly	Phe	Gln	Ser	Asn	Glu	Ala	Ala	Ser	Leu	Ala	Ser
305				310						315					320
Thr	Gly	Val	Gly	Val	Val	Lys	Val	Ile	Ser	Thr	Ile	Pro	Ala	Thr	Leu
				325					330					335	
Leu	Val	Asp	His	Val	Gly	Ser	Lys	Thr	Phe	Leu	Cys	Ile	Gly	Ser	Ser
		340						345					350		
Val	Met	Ala	Ala	Ser	Leu	Val	Thr	Met	Gly	Ile	Val	Asn	Leu	Asn	Ile
	355						360					365			
His	Met	Asn	Phe	Thr	His	Ile	Cys	Arg	Ser	His	Asn	Ser	Ile	Asn	Gln
	370					375					380				
Ser	Leu	Asp	Glu	Ser	Val	Ile	Tyr	Gly	Pro	Gly	Asn	Leu	Ser	Thr	Asn
385				390						395					400
Asn	Asn	Thr	Leu	Arg	Asp	His	Phe	Lys	Gly	Ile	Ser	Ser	His	Ser	Arg
			405						410					415	
Ser	Ser	Leu	Met	Pro	Leu	Arg	Asn	Asp	Val	Asp	Lys	Arg	Gly	Glu	Thr
		420						425					430		
Thr	Ser	Ala	Ser	Leu	Leu	Asn	Ala	Gly	Leu	Ser	His	Thr	Glu	Tyr	Gln
	435						440						445		
Ile	Val	Thr	Asp	Pro	Gly	Asp	Val	Pro	Ala	Phe	Leu	Lys	Trp	Leu	Ser
	450					455					460				
Leu	Ala	Ser	Leu	Leu	Val	Tyr	Val	Ala	Ala	Phe	Ser	Ile	Gly	Leu	Gly
465					470					475					480
Pro	Met	Pro	Trp	Leu	Val	Leu	Ser	Glu	Ile	Phe	Pro	Gly	Gly	Ile	Arg
			485						490					495	
Gly	Arg	Ala	Met	Ala	Leu	Thr	Ser	Ser	Met	Asn	Trp	Gly	Ile	Asn	Leu
		500						505					510		
Leu	Ile	Ser	Leu	Thr	Phe	Leu	Thr	Val	Thr	Asp	Leu	Ile	Gly	Leu	Pro
	515						520					525			
Trp	Val	Cys	Phe	Ile	Tyr	Thr	Ile	Met	Ser	Leu	Ala	Ser	Leu	Leu	Phe
	530					535					540				
Val	Val	Met	Phe	Ile	Pro	Glu	Thr	Lys	Gly	Cys	Ser	Leu	Glu	Gln	Ile
545					550					555					560
Ser	Met	Glu	Leu	Ala	Lys	Val	Asn	Tyr	Val	Lys	Asn	Asn	Ile	Cys	Phe
			565						570				575		
Met	Ser	His	His	Gln	Glu	Glu	Leu	Val	Pro	Lys	Gln	Pro	Gln	Lys	Arg
		580						585				590			
Lys	Pro	Gln	Glu	Gln	Leu	Leu	Glu	Cys	Asn	Lys	Leu	Cys	Gly	Arg	Gly
	595						600					605			
Gln	Ser	Arg	Gln	Leu	Ser	Pro	Glu	Thr							
	610					615									

<210> 3

<211> 1854

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(1854)

<400> 3

atg gta cct gtt gaa aac acc gag ggc ccc agt ctg ctg aac cag aag	48
Met Val Pro Val Glu Asn Thr Glu Gly Pro Ser Leu Leu Asn Gln Lys	
1 5 10 15	
ggg aca gcc gtg gag acg gag ggc agc ggc agc cgg cat cct ccc tgg	96
Gly Thr Ala Val Glu Thr Glu Gly Ser Gly Ser Arg His Pro Pro Trp	
20 25 30	
gcg aga ggc tgc ggc atg ttt acc ttc ctg tca tct gtc act gct gct	144
Ala Arg Gly Cys Gly Met Phe Thr Phe Leu Ser Ser Val Thr Ala Ala	
35 40 45	
gtc agt ggc ctc ctg gtg ggt tat gaa ctt ggg atc atc tct ggg gct	192
Val Ser Gly Leu Leu Val Gly Tyr Glu Leu Gly Ile Ile Ser Gly Ala	
50 55 60	
ctt ctt cag atc aaa acc tta tta gcc ctg agc tgc cat gag cag gaa	240
Leu Leu Gln Ile Lys Thr Leu Leu Ala Leu Ser Cys His Glu Gln Glu	
65 70 75 80	
atg gtt gtg agc tcc ctc gtc att gga gcc ctc ctt gcc tca ctc acc	288
Met Val Val Ser Ser Leu Val Ile Gly Ala Leu Leu Ala Ser Leu Thr	
85 90 95	
gga ggg gtc ctg ata gac aga tat gga aga agg aca gca atc atc ttg	336
Gly Gly Val Leu Ile Asp Arg Tyr Gly Arg Arg Thr Ala Ile Ile Leu	
100 105 110	
tca tcc tgc ctg ctt gga ctc gga agc tta gtc ttg atc ctc agt tta	384
Ser Ser Cys Leu Leu Gly Leu Gly Ser Leu Val Leu Ile Leu Ser Leu	
115 120 125	
tcc tac acg gtt ctt ata gtg gga cgc att gcc ata ggg gtc tcc atc	432
Ser Tyr Thr Val Leu Ile Val Gly Arg Ile Ala Ile Gly Val Ser Ile	
130 135 140	
tcc ctc tct tcc att gcc act tgt gtt tac atc gca gag att gct cct	480
Ser Leu Ser Ser Ile Ala Thr Cys Val Tyr Ile Ala Glu Ile Ala Pro	
145 150 155 160	
caa cac aga aga ggc ctt ctt gtg tca ctg aat gag ctg atg att gtc	528
Gln His Arg Arg Gly Leu Leu Val Ser Leu Asn Glu Leu Met Ile Val	
165 170 175	
atc ggc att ctt tct gcc tat att tca aat tac gca ttt gcc aat gtt	576
Ile Gly Ile Leu Ser Ala Tyr Ile Ser Asn Tyr Ala Phe Ala Asn Val	
180 185 190	
ttc cat ggc tgg aag tac atg ttt ggt ctt gtg att ccc ttg gga gtt	624
Phe His Gly Trp Lys Tyr Met Phe Gly Leu Val Ile Pro Leu Gly Val	
195 200 205	
ttg caa gca att gca atg tat ttt ctt cct cca agc cct cgg ttt ctg	672
Leu Gln Ala Ile Ala Met Tyr Phe Leu Pro Pro Ser Pro Arg Phe Leu	
210 215 220	
gtg atg aaa gga caa gag gga gct gct agc aag gtt ctt gga agg tta	720
Val Met Lys Gly Gln Glu Gly Ala Ala Ser Lys Val Leu Gly Arg Leu	
225 230 235 240	

aga gca ctc tca gat aca act gag gaa ctc act gtg atc aaa tcc tcc	768
Arg Ala Leu Ser Asp Thr Thr Glu Glu Leu Thr Val Ile Lys Ser Ser	
245 250 255	
ctg aaa gat gaa tat cag tac agt ttt tgg gat ctg ttt cgt tca aaa	816
Leu Lys Asp Glu Tyr Gln Tyr Ser Phe Trp Asp Leu Phe Arg Ser Lys	
260 265 270	
gac aac atg cgg acc cga ata atg ata gga cta aca cta gta ttt ttt	864
Asp Asn Met Arg Thr Arg Ile Met Ile Gly Leu Thr Leu Val Phe Phe	
275 280 285	
gta caa atc act ggc caa cca aac ata ttg ttc tat gca tca act gtt	912
Val Gln Ile Thr Gly Gln Pro Asn Ile Leu Phe Tyr Ala Ser Thr Val	
290 295 300	
ttg aag tca gtt gga ttt caa agc aat gag gca gct agc ctc gcc tcc	960
Leu Lys Ser Val Gly Phe Gln Ser Asn Glu Ala Ala Ser Leu Ala Ser	
305 310 315 320	
act ggg gtt gga gtc gtc aag gtc att agc acc atc cct gcc act ctt	1008
Thr Gly Val Gly Val Val Lys Val Ile Ser Thr Ile Pro Ala Thr Leu	
325 330 335	
ctt gta gac cat gtc ggc agc aaa aca ttc ctc tgc att ggc tcc tct	1056
Leu Val Asp His Val Gly Ser Lys Thr Phe Leu Cys Ile Gly Ser Ser	
340 345 350	
gtg atg gca gct tcg ttg gtg acc atg ggc atc gta aat ctc aac atc	1104
Val Met Ala Ala Ser Leu Val Thr Met Gly Ile Val Asn Leu Asn Ile	
355 360 365	
cac atg aac ttc acc cat atc tgc aga agc cac aat tct atc aac cag	1152
His Met Asn Phe Thr His Ile Cys Arg Ser His Asn Ser Ile Asn Gln	
370 375 380	
tcc ttg gat gag tct gtg att tat gga cca gga aac ctg tca acc aac	1200
Ser Leu Asp Glu Ser Val Ile Tyr Gly Pro Gly Asn Leu Ser Thr Asn	
385 390 395 400	
aac aat act ctc aga gac cac ttc aaa ggg att tct tcc cat agc aga	1248
Asn Asn Thr Leu Arg Asp His Phe Lys Gly Ile Ser Ser His Ser Arg	
405 410 415	
agc tca ctc atg ccc ctg aga aat gat gtg gat aag aga ggg gag acg	1296
Ser Ser Leu Met Pro Leu Arg Asn Asp Val Asp Lys Arg Gly Glu Thr	
420 425 430	
acc tca gca tcc ttg cta aat gct gga tta agc cac act gaa tac cag	1344
Thr Ser Ala Ser Leu Leu Asn Ala Gly Leu Ser His Thr Glu Tyr Gln	
435 440 445	
ata gtc aca gac cct ggg gac gtc cca gct ttt ttg aaa tgg ctg tcc	1392
Ile Val Thr Asp Pro Gly Asp Val Pro Ala Phe Leu Lys Trp Leu Ser	
450 455 460	
tta gcc agc ttg ctt gtt tat gtt gct gct ttt tca att ggt cta gga	1440
Leu Ala Ser Leu Leu Val Tyr Val Ala Ala Phe Ser Ile Gly Leu Gly	
465 470 475 480	

cca atg ccc tgg ctg gtg ctc agc gag atc ttt cct ggt ggg atc aga 1488
 Pro Met Pro Trp Leu Val Leu Ser Glu Ile Phe Pro Gly Gly Ile Arg
 485 490 495

gga cga gcc atg gct tta act tct agc atg aac tgg ggc atc aat ctc 1536
 Gly Arg Ala Met Ala Leu Thr Ser Ser Met Asn Trp Gly Ile Asn Leu
 500 505 510

ctc atc tcg ctg aca ttt ttg act gta act gat ctt att ggc ctg cca 1584
 Leu Ile Ser Leu Thr Phe Leu Thr Val Thr Asp Leu Ile Gly Leu Pro
 515 520 525

tgg gtg tgc ttt ata tat aca atc atg agt cta gca tcc ctg ctt ttt 1632
 Trp Val Cys Phe Ile Tyr Thr Ile Met Ser Leu Ala Ser Leu Leu Phe
 530 535 540

gtt gtt atg ttt ata cct gag aca aag gga tgc tct ttg gaa caa ata 1680
 Val Val Met Phe Ile Pro Glu Thr Lys Gly Cys Ser Leu Glu Gln Ile
 545 550 555 560

tca atg gag cta gca aaa gtg aac tat gtg aaa aac aac att tgt ttt 1728
 Ser Met Glu Leu Ala Lys Val Asn Tyr Val Lys Asn Asn Ile Cys Phe
 565 570 575

atg agt cat cac caa gaa gaa tta gtg cca aaa cag cct caa aaa aga 1776
 Met Ser His His Gln Glu Glu Leu Val Pro Lys Gln Pro Gln Lys Arg
 580 585 590

aaa ccc cag gag cag ctc ttg gag tgt aac aag ctg tgt ggt agg ggc 1824
 Lys Pro Gln Glu Gln Leu Leu Glu Cys Asn Lys Leu Cys Gly Arg Gly
 595 600 605

caa tcc agg cag ctt tct cca gag acc taa 1854
 Gln Ser Arg Gln Leu Ser Pro Glu Thr
 610 615

<210> 4
 <211> 2230
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (376)...(1962)

<400> 4
 gtcgacccac gcgtccggca acatggcggc tgccgtggtg cagcgcccg gctgagcgac 60
 agcaagtgca gcgggtcctt accccgggtg aggggtggcc tccgctggg atcgtgccct 120
 cttcagcccg ctctgtccc cgacatcacg tgtattccgc acgtcccctc cgcgtgtgt 180
 gtctactgag acggggaggc gtgacagggc ccgggtccct tctcagtgtt gctctgtgt 240
 tcagggaag ctcccgtct ccgggcgcac ttccctcgcc tgtgttcggt ccatactct 300
 ttctccagcc tctcccctc gcaggtgga tcgtcggtgg gaccggagcg cgggcgggcg 360
 cggcccccg ggacc atg gcc ggg tcc gac acc gcg ccc ttc ctc agc cag 411
 Met Ala Gly Ser Asp Thr Ala Pro Phe Leu Ser Gln
 1 5 10

gcg gat gac ccg gac gac ggg cca gtg cct ggc acc ccg ggg ttg cca 459
 Ala Asp Asp Pro Asp Asp Gly Pro Val Pro Gly Thr Pro Gly Leu Pro
 15 20 25

ggg tcc acg ggg aac ccg aag tcc gag gag ccc gag gtc ccg gac cag	507
Gly Ser Thr Gly Asn Pro Lys Ser Glu Glu Pro Glu Val Pro Asp Gln	
30 35 40	
gag ggg ctg cag cgc atc acc ggc ctg tct ccc ggc cgt tgc gct ctc	555
Glu Gly Leu Gln Arg Ile Thr Gly Leu Ser Pro Gly Arg Ser Ala Leu	
45 50 55 60	
ata gtg gcg gtg ctg tgc tac atc aat ctc ctg aac tac atg gac cgc	603
Ile Val Ala Val Leu Cys Tyr Ile Asn Leu Leu Asn Tyr Met Asp Arg	
65 70 75	
ttc acc gtg gct ggc gtc ctt ccc gac atc gag cag ttc ttc aac atc	651
Phe Thr Val Ala Gly Val Leu Pro Asp Ile Glu Gln Phe Phe Asn Ile	
80 85 90	
ggg gac agt agc tct ggg ctc atc cag acc gtg ttc atc tcc agt tac	699
Gly Asp Ser Ser Ser Gly Leu Ile Gln Thr Val Phe Ile Ser Ser Tyr	
95 100 105	
atg gtg ttg gca cct gtg ttt ggc tac ctg ggt gac agg tac aat cgg	747
Met Val Leu Ala Pro Val Phe Gly Tyr Leu Gly Asp Arg Tyr Asn Arg	
110 115 120	
aag tat ctc atg tgc ggg ggc att gcc ttc tgg tcc ctg gtg aca ctg	795
Lys Tyr Leu Met Cys Gly Gly Ile Ala Phe Trp Ser Leu Val Thr Leu	
125 130 135 140	
ggg tca tcc ttc atc ccc gga gag cat ttc tgg ctg ctc ctc ctg acc	843
Gly Ser Ser Phe Ile Pro Gly Glu His Phe Trp Leu Leu Leu Leu Thr	
145 150 155	
cgg ggc ctg gtg ggg gtc ggg gag gcc agt tat tcc acc atc gcg ccc	891
Arg Gly Leu Val Gly Val Gly Glu Ala Ser Tyr Ser Thr Ile Ala Pro	
160 165 170	
act ctc att gcc gac ctc ttt gtg gcc gac cag cgg agc cgg atg ctc	939
Thr Leu Ile Ala Asp Leu Phe Val Ala Asp Gln Arg Ser Arg Met Leu	
175 180 185	
agc atc ttc tac ttt gcc att ccg gtg ggc agt ggt ctg ggc tac att	987
Ser Ile Phe Tyr Phe Ala Ile Pro Val Gly Ser Gly Leu Gly Tyr Ile	
190 195 200	
gca ggc tcc aaa gtg aag gat atg gct gga gac tgg cac tgg gct ctg	1035
Ala Gly Ser Lys Val Lys Asp Met Ala Gly Asp Trp His Trp Ala Leu	
205 210 215 220	
agg gtg aca ccg ggt cta gga gtg gtg gcc gtt ctg ctg ctg ttc ctg	1083
Arg Val Thr Pro Gly Leu Gly Val Val Ala Val Leu Leu Leu Phe Leu	
225 230 235	
gta gtg cgg gag ccg cca agg gga gcc gtg gag cgc cac tca gat ttg	1131
Val Val Arg Glu Pro Pro Arg Gly Ala Val Glu Arg His Ser Asp Leu	
240 245 250	
cca ccc ctg aac ccc acc tgc tgg tgg gca gat ctg agg gct ctg gca	1179
Pro Pro Leu Asn Pro Thr Ser Trp Trp Ala Asp Leu Arg Ala Leu Ala	
255 260 265	

aga aat cct agt ttc gtc ctg tct tcc ctg ggc ttc act gct gtg gcc	1227
Arg Asn Pro Ser Phe Val Leu Ser Ser Leu Gly Phe Thr Ala Val Ala	
270 275 280	
ttt gtc acg ggc tcc ctg gct ctg tgg gct ccg gca ttc ctg ctg cgt	1275
Phe Val Thr Gly Ser Leu Ala Leu Trp Ala Pro Ala Phe Leu Leu Arg	
285 290 295 300	
tcc cgc gtg gtc ctt ggg gag acc cca ccc tgc ctt ccc gga gac tcc	1323
Ser Arg Val Val Leu Gly Glu Thr Pro Cys Leu Pro Gly Asp Ser	
305 310 315	
tgc tct tcc tct gac agt ctc atc ttt gga ctc atc acc tgc ctg acc	1371
Cys Ser Ser Ser Asp Ser Leu Ile Phe Gly Leu Ile Thr Cys Leu Thr	
320 325 330	
gga gtc ctg ggt gtg ggc ctg ggt gtg gag atc agc cgc cgg ctc cgc	1419
Gly Val Leu Gly Val Gly Leu Gly Val Glu Ile Ser Arg Arg Leu Arg	
335 340 345	
cac tcc aac ccc cgg gct gat ccc ctg gtc tgt gcc act ggc ctc ctg	1467
His Ser Asn Pro Arg Ala Asp Pro Leu Val Cys Ala Thr Gly Leu Leu	
350 355 360	
ggc tct gca ccc ttc ctc ttc ctg tcc ctt gcc tgc gcc cgt ggt agc	1515
Gly Ser Ala Pro Phe Leu Phe Leu Ser Leu Ala Cys Ala Arg Gly Ser	
365 370 375 380	
atc gtg gcc act tat att ttc atc ttc att gga gag acc ctc ctg tcc	1563
Ile Val Ala Thr Tyr Ile Phe Ile Phe Ile Gly Glu Thr Leu Leu Ser	
385 390 395	
atg aac tgg gcc atc gtg gcc gac att ctg ctg tac gtg gtg atc cct	1611
Met Asn Trp Ala Ile Val Ala Asp Ile Leu Leu Tyr Val Val Ile Pro	
400 405 410	
acc cga cgc tcc acc gcc gag gcc ttc cag atc gtg ctg tcc cac ctg	1659
Thr Arg Arg Ser Thr Ala Glu Ala Phe Gln Ile Val Leu Ser His Leu	
415 420 425	
ctg ggt gat gct ggg agc ccc tac ctc att ggc ctg atc tct gac cgc	1707
Leu Gly Asp Ala Gly Ser Pro Tyr Leu Ile Gly Leu Ile Ser Asp Arg	
430 435 440	
ctg cgc cgg aac tgg ccc ccc tcc ttc ttg tcc gag ttc cgg gct ctg	1755
Leu Arg Arg Asn Trp Pro Pro Ser Phe Leu Ser Glu Phe Arg Ala Leu	
445 450 455 460	
cag ttc tcg ctc atg ctc tgc gcg ttt gtt ggg gca ctg ggc ggc gca	1803
Gln Phe Ser Leu Met Leu Cys Ala Phe Val Gly Ala Leu Gly Gly Ala	
465 470 475	
gcc ttc ctg ggc acc gcc atc ttc att gag gcc gac cgc cgg cgg gca	1851
Ala Phe Leu Gly Thr Ala Ile Phe Ile Glu Ala Asp Arg Arg Arg Ala	
480 485 490	
cag ctg cac gtg cag ggc ctg ctg cac gaa gca ggg tcc aca gac gac	1899
Gln Leu His Val Gln Gly Leu Leu His Glu Ala Gly Ser Thr Asp Asp	
495 500 505	

cgg att gtg gtg ccc cag cgg ggc cgc tcc acc cgc gtg ccc gtg gcc 1947
 Arg Ile Val Val Pro Gln Arg Gly Arg Ser Thr Arg Val Pro Val Ala
 510 " 515 " 520

agt gtg ctc atc tga gaggtgtgccc ctcacctacc tgcacatctg ccacagctgg 2002
 Ser Val Leu Ile *
 525

ccctggggccc accccacgaa gggcctgggc ctaacccctt ggcttgccc agcttccaga 2062
 gggaccctgg gccgtgtgcc agctccaga cactacatgg gtagctcagg ggaggaggtg 2122
 ggggtccagg agggggatcc ctctccacag gggcagcccc aagggtcgg tgctatttgt 2182
 aacggaataa aattttagc cagaaaaaaa aaaaaaagg' gcggccgc 2230

<210> 5
 <211> 528
 <212> PRT
 <213> Homo sapiens

<400> 5
 Met Ala Gly Ser Asp Thr Ala Pro Phe Leu Ser Gln Ala Asp Asp Pro
 1 5 10 15
 Asp Asp Gly Pro Val Pro Gly Thr Pro Gly Leu Pro Gly Ser Thr Gly
 20 25 30
 Asn Pro Lys Ser Glu Glu Pro Glu Val Pro Asp Gln Glu Gly Leu Gln
 35 40 45
 Arg Ile Thr Gly Leu Ser Pro Gly Arg Ser Ala Leu Ile Val Ala Val
 50 55 60
 Leu Cys Tyr Ile Asn Leu Asn Tyr Met Asp Arg Phe Thr Val Ala
 65 70 75 80
 Gly Val Leu Pro Asp Ile Glu Gln Phe Phe Asn Ile Gly Asp Ser Ser
 85 90 95
 Ser Gly Leu Ile Gln Thr Val Phe Ile Ser Ser Tyr Met Val Leu Ala
 100 105 110
 Pro Val Phe Gly Tyr Leu Gly Asp Arg Tyr Asn Arg Lys Tyr Leu Met
 115 120 125
 Cys Gly Gly Ile Ala Phe Trp Ser Leu Val Thr Leu Gly Ser Ser Phe
 130 135 140
 Ile Pro Gly Glu His Phe Trp Leu Leu Leu Thr Arg Gly Leu Val
 145 150 155 160
 Gly Val Gly Glu Ala Ser Tyr Ser Thr Ile Ala Pro Thr Leu Ile Ala
 165 170 175
 Asp Leu Phe Val Ala Asp Gln Arg Ser Arg Met Leu Ser Ile Phe Tyr
 180 185 190
 Phe Ala Ile Pro Val Gly Ser Gly Leu Gly Tyr Ile Ala Gly Ser Lys
 195 200 205
 Val Lys Asp Met Ala Gly Asp Trp His Trp Ala Leu Arg Val Thr Pro
 210 215 220
 Gly Leu Gly Val Val Ala Val Leu Leu Leu Phe Leu Val Val Arg Glu
 225 230 235 240
 Pro Pro Arg Gly Ala Val Glu Arg His Ser Asp Leu Pro Pro Leu Asn
 245 250 255
 Pro Thr Ser Trp Trp Ala Asp Leu Arg Ala Leu Ala Arg Asn Pro Ser
 260 265 270
 Phe Val Leu Ser Ser Leu Gly Phe Thr Ala Val Ala Phe Val Thr Gly
 275 280 285
 Ser Leu Ala Leu Trp Ala Pro Ala Phe Leu Leu Arg Ser Arg Val Val
 290 295 300
 Leu Gly Glu Thr Pro Pro Cys Leu Pro Gly Asp Ser Cys Ser Ser Ser
 305 310 315 320
 Asp Ser Leu Ile Phe Gly Leu Ile Thr Cys Leu Thr Gly Val Leu Gly
 325 330 335

Val Gly Leu Gly Val Glu Ile Ser Arg Arg Leu Arg His Ser Asn Pro
 340 345 350
 Arg Ala Asp Pro Leu Val Cys Ala Thr Gly Leu Leu Gly Ser Ala Pro
 355 360 365
 Phe Leu Phe Leu Ser Leu Ala Cys Ala Arg Gly Ser Ile Val Ala Thr
 370 375 380
 Tyr Ile Phe Ile Phe Ile Gly Glu Thr Leu Leu Ser Met Asn Trp Ala
 385 390 395 400
 Ile Val Ala Asp Ile Leu Leu Tyr Val Val Ile Pro Thr Arg Arg Ser
 405 410 415
 Thr Ala Glu Ala Phe Gln Ile Val Leu Ser His Leu Leu Gly Asp Ala
 420 425 430
 Gly Ser Pro Tyr Leu Ile Gly Leu Ile Ser Asp Arg Leu Arg Arg Asn
 435 440 445
 Trp Pro Pro Ser Phe Leu Ser Glu Phe Arg Ala Leu Gln Phe Ser Leu
 450 455 460
 Met Leu Cys Ala Phe Val Gly Ala Leu Gly Gly Ala Ala Phe Leu Gly
 465 470 475 480
 Thr Ala Ile Phe Ile Glu Ala Asp Arg Arg Ala Gln Leu His Val
 485 490 495
 Gln Gly Leu Leu His Glu Ala Gly Ser Thr Asp Asp Arg Ile Val Val
 500 505 510
 Pro Gln Arg Gly Arg Ser Thr Arg Val Pro Val Ala Ser Val Leu Ile
 515 520 525

<210> 6

<211> 1587

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(1587)

<400> 6

atg gcc ggg tcc gac acc gcg ccc ttc ctc agc cag gcg gat gac ccg 48
 Met Ala Gly Ser Asp Thr Ala Pro Phe Leu Ser Gln Ala Asp Pro
 1 5 10 15
 gac gac ggg cca gtg cct ggc acc ccg ggg ttg cca ggg tcc acg ggg 96
 Asp Asp Gly Pro Val Pro Gly Thr Pro Gly Leu Pro Gly Ser Thr Gly
 20 25 30
 aac ccg aag tcc gag gag ccc gag gtc ccg gac cag gag ggg ctg cag 144
 Asn Pro Lys Ser Glu Glu Pro Glu Val Pro Asp Gln Glu Gly Leu Gln
 35 40 45
 cgc atc acc ggc ctg tct ccc ggc cgt tcg gct ctc ata gtg gcg gtg 192
 Arg Ile Thr Gly Leu Ser Pro Gly Arg Ser Ala Leu Ile Val Ala Val
 50 55 60
 ctg tgc tac atc aat ctc ctg aac tac atg gac cgc ttc acc gtg gct 240
 Leu, Cys Tyr Ile Asn Leu Leu Asn Tyr Met Asp Arg Phe Thr Val Ala
 65 70 75 80
 ggc gtc ctt ccc gac atc gag cag ttc ttc aac atc ggg gac agt agc 288
 Gly Val Leu Pro Asp Ile Glu Gln Phe Asn Ile Gly Asp Ser Ser
 85 90 95

tct ggg ctc atc cag acc gtg ttc atc tcc agt tac atg gtg ttg gca	336
Ser Gly Leu Ile Gln Thr Val Phe Ile Ser Ser Tyr Met Val Leu Ala	
100 105 110	
cct gtg ttt ggc tac ctg ggt gac agg tac aat cgg aag tat ctc atg	384
Pro Val Phe Gly Tyr Leu Gly Asp Arg Tyr Asn Arg Lys Tyr Leu Met	
115 120 125	
tgc ggg ggc att gcc ttc tgg tcc ctg gtg aca ctg ggg tca tcc ttc	432
Cys Gly Gly Ile Ala Phe Trp Ser Leu Val Thr Leu Gly Ser Ser Phe	
130 135 140	
atc ccc gga gag cat ttc tgg ctg ctc ctc ctg acc cgg ggc ctg gtg	480
Ile Pro Gly Glu His Phe Trp Leu Leu Leu Leu Thr Arg Gly Leu Val	
145 150 155 160	
ggg gtc ggg gag gcc agt tat tcc acc atc gcg ccc act ctc att gcc	528
Gly Val Gly Glu Ala Ser Tyr Ser Thr Ile Ala Pro Thr Leu Ile Ala	
165 170 175	
gac ctc ttt gtg gcc gac cag cgg agc cgg atg ctc agc atc ttc tac	576
Asp Leu Phe Val Ala Asp Gln Arg Ser Arg Met Leu Ser Ile Phe Tyr	
180 185 190	
ttt gcc att ccg gtg ggc agt ggt ctg ggc tac att gca ggc tcc aaa	624
Phe Ala Ile Pro Val Gly Ser Gly Leu Gly Tyr Ile Ala Gly Ser Lys	
195 200 205	
gtg aag gat atg gct gga gac tgg cac tgg gct ctg agg gtg aca ccg	672
Val Lys Asp Met Ala Gly Asp Trp His Trp Ala Leu Arg Val Thr Pro	
210 215 220	
ggt cta gga gtg gtg gcc gtt ctg ctg ctg ttc ctg gta gtg cgg gag	720
Gly Leu Gly Val Val Ala Val Leu Leu Leu Phe Leu Val Val Arg Glu	
225 230 235 240	
ccg cca agg gga gcc gtg gag cgc cac tca gat ttg cca ccc ctg aac	768
Pro Pro Arg Gly Ala Val Glu Arg His Ser Asp Leu Pro Pro Leu Asn	
245 250 255	
ccc acc tcg tgg tgg gca gat ctg agg gct ctg gca aga aat cct agt	816
Pro Thr Ser Trp Trp Ala Asp Leu Arg Ala Leu Ala Arg Asn Pro Ser	
260 265 270	
ttc gtc ctg tct tcc ctg ggc ttc act gct gtg gcc ttt gtc acg ggc	864
Phe Val Leu Ser Ser Leu Gly Phe Thr Ala Val Ala Phe Val Thr Gly	
275 280 285	
tcc ctg gct ctg tgg gct ccg gca ttc ctg ctg cgt tcc cgc gtg gtc	912
Ser Leu Ala Leu Trp Ala Pro Ala Phe Leu Leu Arg Ser Arg Val Val	
290 295 300	
ctt ggg gag acc cca ccc tgc ctt ccc gga gac tcc tgc tct tcc tct	960
Leu Gly Glu Thr Pro Pro Cys Leu Pro Gly Asp Ser Cys Ser Ser Ser	
305 310 315 320	
gac agt ctc atc ttt gga ctc atc acc tgc ctg acc gga gtc ctg ggt	1008
Asp Ser Leu Ile Phe Gly Leu Ile Thr Cys Leu Thr Gly Val Leu Gly	
325 330 335	

gtg ggc ctg ggt gtg gag atc agc cgc cgg ctc cgc cac tcc aac ccc	1056
Val Gly Leu Gly Val Glu Ile Ser Arg Arg Leu Arg His Ser Asn Pro	
340 345 350	
cgg gct gat ccc ctg gtc tgt gcc act ggc ctc ctg ggc tct gca ccc	1104
Arg Ala Asp Pro Leu Val Cys Ala Thr Gly Leu Leu Gly Ser Ala Pro	
355 360 365	
ttc ctc ttc ctg tcc ctt gcc tgc gcc cgt ggt agc atc gtg gcc act	1152
Phe Leu Phe Leu Ser Leu Ala Cys Ala Arg Gly Ser Ile Val Ala Thr	
370 375 380	
tat att ttc atc ttc att gga gag acc ctc ctg tcc atg aac tgg gcc	1200
Tyr Ile Phe Ile Phe Ile Gly Glu Thr Leu Leu Ser Met Asn Trp Ala	
385 390 395 400	
atc gtg gcc gac att ctg ctg tac gtg gtg atc cct acc cga cgc tcc	1248
Ile Val Ala Asp Ile Leu Leu Tyr Val Val Ile Pro Thr Arg Arg Ser	
405 410 415	
acc gcc gag gcc ttc cag atc gtg ctg tcc cac ctg ctg ggt gat gct	1296
Thr Ala Glu Ala Phe Gln Ile Val Leu Ser His Leu Leu Gly Asp Ala	
420 425 430	
ggg agc ccc tac ctc att ggc ctg atc tct gac cgc ctg cgc cgg aac	1344
Gly Ser Pro Tyr Leu Ile Gly Leu Ile Ser Asp Arg Leu Arg Arg Asn	
435 440 445	
tgg ccc ccc tcc ttc ttg tcc gag ttc cgg gct ctg cag ttc tcg ctc	1392
Trp Pro Pro Ser Phe Leu Ser Glu Phe Arg Ala Leu Gln Phe Ser Leu	
450 455 460	
atg ctc tgc gcg ttt gtt ggg gca ctg ggc ggc gca gcc ttc ctg ggc	1440
Met Leu Cys Ala Phe Val Gly Ala Leu Gly Gly Ala Ala Phe Leu Gly	
465 470 475 480	
acc gcc atc ttc att gag gcc gac cgc cgg cgg gca cag ctg cac gtg	1488
Thr Ala Ile Phe Ile Glu Ala Asp Arg Arg Arg Ala Gln Leu His Val	
485 490 495	
cag ggc ctg ctg cac gaa gca ggg tcc aca gac gac cgg att gtg gtg	1536
Gln Gly Leu Leu His Glu Ala Gly Ser Thr Asp Asp Arg Ile Val Val	
500 505 510	
ccc cag cgg ggc cgc tcc acc cgc gtg ccc gtg gcc agt gtg ctc atc	1584
Pro Gln Arg Gly Arg Ser Thr Arg Val Pro Val Ala Ser Val Leu Ile	
515 520 525	
tga	1587
*	

<210> 7

<211> 4632

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (225) ... (3581)

<220>

<221> misc_feature

<222> (4611)

<223> n = a, c, t, or g

<400> 7

cacgcgtccg cccacgcgtc cgcccacgcg tccgagcccc ctttcaagcc ttagcttccg 60
 gctccaagcc gacccctccc cctccctgt ccccttcccc ttctcccatc cctctctcgg 120
 ccacagcgtc ttgttagtcc tctccctcta ctccgcaata tttctttct ttctccctcc 180
 tctccctcat ttgttgtttg atgtttccca ctctttgagg aagg atg gtt gat ttg 236
 Met Val Asp Leu

gag agc gaa gtg ccc cct ctg cct ccc agg tac agg ttt cga gat ttg 284
 Glu Ser Glu Val Pro Pro Leu Pro Pro Arg Tyr Arg Phe Arg Asp Leu
 5 10 15 20

ctg cta ggg gac caa gga tgg caa aac gat gac aga gta caa gtt gaa 332
 Leu Leu Gly Asp Gln Gly Trp Gln Asn Asp Arg Val Gln Val Glu
 25 30 35

ttc tat atg aat gaa aat aca ttt aaa gaa aga cta aaa tta ttt ttc 380
 Phe Tyr Met Asn Glu Asn Thr Phe Lys Glu Arg Leu Lys Leu Phe Phe
 40 45 50

ata aaa aac cag aga tca agt cta agg ata cgc ctg ttc aat ttt tct 428
 Ile Lys Asn Gln Arg Ser Ser Leu Arg Ile Arg Leu Phe Asn Phe Ser
 55 60 65

ctc aaa tta cta agc tgc tta tta tac ata atc cga gta cta cta gaa 476
 Leu Lys Leu Leu Ser Cys Leu Leu Tyr Ile Ile Arg Val Leu Leu Glu
 70 75 80

aac cct tca caa gga aat gaa tgg tct cat atc ttt tgg gtg aac aga 524
 Asn Pro Ser Gln Gly Asn Glu Trp Ser His Ile Phe Trp Val Asn Arg
 85 90 95 100

agt cta cct ttg tgg ggc tta cag gtt tca gtg gca ttg ata agt ctg 572
 Ser Leu Pro Leu Trp Gly Leu Gln Val Ser Val Ala Leu Ile Ser Leu
 105 110 115

ttt gaa aca ata tta ctt ggt tat ctt agt tat aag gga aac atc tgg 620
 Phe Glu Thr Ile Leu Leu Gly Tyr Leu Ser Tyr Lys Gly Asn Ile Trp
 120 125 130

gaa cag att tta cga ata ccc ttc atc ttg gaa ata att aat gca gtt 668
 Glu Gln Ile Leu Arg Ile Pro Phe Ile Leu Glu Ile Ile Asn Ala Val
 135 140 145

ccc ttc att atc tca ata ttc tgg cct tcc tta agg aat cta ttt gtc 716
 Pro Phe Ile Ile Ser Ile Phe Trp Pro Ser Leu Arg Asn Leu Phe Val
 150 155 160

cca gtc ttt ctg aac tgt tgg ctt gcc aaa cat gcc ttg gaa aat atg 764
 Pro Val Phe Leu Asn Cys Trp Leu Ala Lys His Ala Leu Glu Asn Met
 165 170 175 180

att aat gat cta cac aga gcc att cag cgt aca cag tct gca atg ttt 812
 Ile Asn Asp Leu His Arg Ala Ile Gln Arg Thr Gln Ser Ala Met Phe
 185 190 195

aat	caa	gtt	ttg	att	tta	ata	tct	aca	tta	cta	tgc	ctt	atc	ttc	acc	860
Asn	Gln	Val	Leu	Ile	Leu	Ile	Ser	Thr	Leu	Leu	Cys	Leu	Ile	Phe	Thr	
		200						205					210			
tgc	att	tgt	ggg	atc	caa	cat	ctg	gaa	cga	ata	gga	aag	agg	ctg	aat	908
Cys	Ile	Cys	Gly	Ile	Gln	His	Leu	Glu	Arg	Ile	Gly	Lys	Arg	Leu	Asn	
		215					220					225				
ctc	ttt	gac	tcc	ctt	tat	ttc	tgc	att	gtg	acg	ttt	tct	act	gtg	ggc	956
Leu	Phe	Asp	Ser	Leu	Tyr	Phe	Cys	Ile	Val	Thr	Phe	Ser	Thr	Val	Gly	
	230					235					240					
ttc	ggg	gat	gtc	act	cct	gaa	aca	tgg	tcc	tcc	aag	ctt	ttt	gta	gtt	1004
Phe	Gly	Asp	Val	Thr	Pro	Glu	Thr	Trp	Ser	Ser	Lys	Leu	Phe	Val	Val	
245					250				255						260	
gct	atg	att	tgt	gtt	gct	ctt	gtg	gtt	cta	ccc	ata	cag	ttt	gaa	cag	1052
Ala	Met	Ile	Cys	Val	Ala	Leu	Val	Val	Leu	Pro	Ile	Gln	Phe	Glu	Gln	
				265					270					275		
ctg	gct	tat	ttg	tgg	atg	gag	aga	caa	aag	tca	gga	gga	aac	tat	agt	1100
Leu	Ala	Tyr	Leu	Trp	Met	Glu	Arg	Gln	Lys	Ser	Gly	Gly	Asn	Tyr	Ser	
		280					285						290			
cga	cat	aga	gct	caa	act	gaa	aag	cat	gtc	gtc	ctg	tgt	gtc	agc	tca	1148
Arg	His	Arg	Ala	Gln	Thr	Glu	Lys	His	Val	Val	Leu	Cys	Val	Ser	Ser	
		295					300					305				
ctg	aag	att	gat	tta	ctt	atg	gat	ttt	tta	aat	gaa	ttc	tat	gct	cat	1196
Leu	Lys	Ile	Asp	Leu	Leu	Met	Asp	Phe	Leu	Asn	Glu	Phe	Tyr	Ala	His	
	310					315					320					
cct	agg	ctc	cag	gat	tat	tat	gtg	gtg	att	ttg	tgt	cct	act	gaa	atg	1244
Pro	Arg	Leu	Gln	Asp	Tyr	Tyr	Val	Val	Ile	Leu	Cys	Pro	Thr	Glu	Met	
325					330					335					340	
gat	gta	cag	gtt	cga	agg	gta	ctg	cag	att	cca	atg	tgg	tcc	caa	cga	1292
Asp	Val	Gln	Val	Arg	Arg	Val	Leu	Gln	Ile	Pro	Met	Trp	Ser	Gln	Arg	
				345					350					355		
gtt	atc	tac	ctt	caa	ggt	tca	gcc	ctt	aaa	gat	caa	gac	cta	ttg	aga	1340
Val	Ile	Tyr	Leu	Gln	Gly	Ser	Ala	Leu	Lys	Asp	Gln	Asp	Leu	Leu	Arg	
			360				365						370			
gca	aag	atg	gat	gac	gct	gag	gcc	tgt	ttt	att	ctc	agt	agc	cgt	tgt	1388
Ala	Lys	Met	Asp	Asp	Ala	Glu	Ala	Cys	Phe	Ile	Leu	Ser	Ser	Arg	Cys	
		375					380					385				
gaa	gtg	gat	agg	aca	tca	tct	gat	cac	caa	aca	att	ttg	aga	gca	tgg	1436
Glu	Val	Asp	Arg	Thr	Ser	Ser	Asp	His	Gln	Thr	Ile	Leu	Arg	Ala	Trp	
	390					395					400					
gct	gtg	aaa	gat	ttt	gct	cca	aat	tgt	cct	ttg	tat	gtc	cag	ata	tta	1484
Ala	Val	Lys	Asp	Phe	Ala	Pro	Asn	Cys	Pro	Leu	Tyr	Val	Gln	Ile	Leu	
405					410				415						420	
aag	cct	gaa	aat	aaa	ttt	cac	atc	aaa	ttt	gct	gat	cat	gtt	gtt	tgt	1532
Lys	Pro	Glu	Asn	Lys	Phe	His	Ile	Lys	Phe	Ala	Asp	His	Val	Val	Cys	
				425					430					435		

gaa gaa gag ttt aaa tac gcc atg tta gct tta aac tgt ata tgc cca	1580
Glu Glu Glu Phe Lys Tyr Ala Met Leu Ala Leu Asn Cys Ile Cys Pro	
440 445 450	
gca aca tct aca ctt att aca cta ctg gtt cat acc tct aga ggg caa	1628
Ala Thr Ser Thr Leu Ile Thr Leu Leu Val His Thr Ser Arg Gly Gln	
455 460 465	
gaa ggc cag caa tcg cca gaa caa tgg cag aag atg tac ggt aga tgc	1676
Glu Gly Gln Gln Ser Pro Glu Gln Trp Gln Lys Met Tyr Gly Arg Cys	
470 475 480	
tcc ggg aat gaa gtc tac cac att gtt ttg gaa gaa agt aca ttt ttt	1724
Ser Gly Asn Glu Val Tyr His Ile Val Leu Glu Glu Ser Thr Phe Phe	
485 490 495 500	
gct gaa tat gaa gga aag agt ttt aca tat gcc tct ttc cat gca cac	1772
Ala Glu Tyr Glu Gly Lys Ser Phe Thr Tyr Ala Ser Phe His Ala His	
505 510 515	
aaa aag ttt ggc gtc tgc ttg att ggt gtt agg agg gag gat aat aaa	1820
Lys Lys Phe Gly Val Cys Leu Ile Gly Val Arg Arg Glu Asp Asn Lys	
520 525 530	
aac att ttg ctg aat cca ggt cct cga tac att atg aat tct acg gac	1868
Asn Ile Leu Leu Asn Pro Gly Pro Arg Tyr Ile Met Asn Ser Thr Asp	
535 540 545	
ata tgc ttt tat att aat att acc aaa gaa gag aat tca gca ttt aaa	1916
Ile Cys Phe Tyr Ile Asn Ile Thr Lys Glu Glu Asn Ser Ala Phe Lys	
550 555 560	
aac caa gac cag cag aga aaa agc aat gtg tcc agg tcg ttt tat cat	1964
Asn Gln Asp Gln Gln Arg Lys Ser Asn Val Ser Arg Ser Phe Tyr His	
565 570 575 580	
gga cct tcc aga tta cct gta cat agc ata att gcc agc atg ggt act	2012
Gly Pro Ser Arg Leu Pro Val His Ser Ile Ile Ala Ser Met Gly Thr	
585 590 595	
gtg gct ata gac ctg caa gat aca agc tgt aga tca gca agt ggc cct	2060
Val Ala Ile Asp Leu Gln Asp Thr Ser Cys Arg Ser Ala Ser Gly Pro	
600 605 610	
acc ctg tct ctt cct aca gag gga agc aaa gaa ata aga aga cct agc	2108
Thr Leu Ser Leu Pro Thr Glu Gly Ser Lys Glu Ile Arg Arg Pro Ser	
615 620 625	
att gct cct gtt tta gag gtt gca gat aca tca tcg att caa aca tgt	2156
Ile Ala Pro Val Leu Glu Val Ala Asp Thr Ser Ser Ile Gln Thr Cys	
630 635 640	
gat ctt cta agt gac caa tca gaa gat gaa act aca cca gat gaa gaa	2204
Asp Leu Leu Ser Asp Gln Ser Glu Asp Glu Thr Thr Pro Asp Glu Glu	
645 650 655 660	
atg tct tca aac tta gag tat gct aaa ggt tac cca cct tat tct cca	2252
Met Ser Ser Asn Leu Glu Tyr Ala Lys Gly Tyr Pro Pro Tyr Ser Pro	
665 670 675	

tat ata gga agt tca ccc act ttt tgt cat ctc ctt cat gaa aaa gta	2300
Tyr Ile Gly Ser Ser Pro Thr Phe Cys His Leu Leu His Glu Lys Val	
680 685 690	
cca ttt tgc tgc tta aga tta gac aag agt tgc caa cat aac tac tat	2348
Pro Phe Cys Cys Leu Arg Leu Asp Lys Ser Cys Gln His Asn Tyr Tyr	
695 700 705	
gag gat gca aaa gcc tat gga ttc aaa aat aaa cta att ata gtt gca	2396
Glu Asp Ala Lys Ala Tyr Gly Phe Lys Asn Lys Leu Ile Ile Val Ala	
710 715 720	
gct gaa aca gct gga aat gga tta tat aac ttt att gtt cct ctc agg	2444
Ala Glu Thr Ala Gly Asn Gly Leu Tyr Asn Phe Ile Val Pro Leu Arg	
725 730 735 740	
gca tat tat aga cca aag aaa gaa ctt aat ccc ata gta ctg cta ttg	2492
Ala Tyr Tyr Arg Pro Lys Lys Glu Leu Asn Pro Ile Val Leu Leu	
745 750 755	
gat aac ccc cta gat gac tta ctc agg tgt gga gtg act ttt gct gct	2540
Asp Asn Pro Leu Asp Asp Leu Leu Arg Cys Gly Val Thr Phe Ala Ala	
760 765 770	
aat atg gtg gtt gtg gat aaa gag agc acc atg agt gcc gag gaa gac	2588
Asn Met Val Val Val Asp Lys Glu Ser Thr Met Ser Ala Glu Glu Asp	
775 780 785	
tac atg gca gat gcc aaa acc att gtg aac gtg cag aca ctc ttc agg	2636
Tyr Met Ala Asp Ala Lys Thr Ile Val Asn Val Gln Thr Leu Phe Arg	
790 795 800	
ttg ttt tcc agt ctc agt att atc aca gag cta act cac ccc gcc aac	2684
Leu Phe Ser Ser Leu Ser Ile Ile Thr Glu Leu Thr His Pro Ala Asn	
805 810 815 820	
atg aga ttc atg caa ttc aga gcc aaa gac tgt tac tct ctt gct ctt	2732
Met Arg Phe Met Gln Phe Arg Ala Lys Asp Cys Tyr Ser Leu Ala Leu	
825 830 835	
tca aaa ctg gaa aag aaa gaa cgg gag aga ggc tct aac ttg gcc ttt	2780
Ser Lys Leu Glu Lys Lys Glu Arg Glu Arg Gly Ser Asn Leu Ala Phe	
840 845 850	
atg ttt tga ctg cct ttt gct gct ggg agg gtg ttt agc atc agt atg	2828
Met Phe Arg Leu Pro Phe Ala Ala Gly Arg Val Phe Ser Ile Ser Met	
855 860 865	
ttg gac act ctg ctg tat cag tca ttt gtg aag gat tat atg att tct	2876
Leu Asp Thr Leu Leu Tyr Gln Ser Phe Val Lys Asp Tyr Met Ile Ser	
870 875 880	
atc acg aga ctt ctg ttg gga ctg gac act aca cca gga tct ggg ttt	2924
Ile Thr Arg Leu Leu Leu Gly Leu Asp Thr Thr Pro Gly Ser Gly Phe	
885 890 895 900	
ctt tgt tct atg aaa atc act gca gat gac tta tgg atc aga act tat	2972
Leu Cys Ser Met Lys Ile Thr Ala Asp Asp Leu Trp Ile Arg Thr Tyr	
905 910 915	

gcc aga ctt tat cag aag ttg tgt tct tct act gga gat gtt ccc att 3020
 Ala Arg Leu Tyr Gln Lys Leu Cys Ser Ser Thr Gly Asp Val Pro Ile
 920 925 930

gga atc tac agg act gag tct cag aaa ctt act aca tct gag tct cga 3068
 Gly Ile Tyr Arg Thr Glu Ser Gln Lys Leu Thr Thr Ser Glu Ser Arg
 935 940 945

aaa ata gca tca caa tct caa ata tct atc agt gta gaa gag tgg gaa 3116
 Lys Ile Ala Ser Gln Ser Gln Ile Ser Ile Ser Val Glu Glu Trp Glu
 950 955 960

gac acc aaa gac tcc aaa gaa caa ggg cac cac cgc agc aac cac cgc 3164
 Asp Thr Lys Asp Ser Lys Glu Gln Gly His His Arg Ser Asn His Arg
 965 970 975 980

aac tca aca tcc agt gac cag tgc gac cat ccc ttg ctg cgg aga aaa 3212
 Asn Ser Thr Ser Ser Asp Gln Ser Asp His Pro Leu Leu Arg Arg Lys
 985 990 995

agc atg cag tgg gcc cga aga ctg agc aga aaa ggc cca aap cac tct 3260
 Ser Met Gln Trp Ala Arg Arg Leu Ser Arg Lys Gly Pro Lys His Ser
 1000 1005 1010

ggt aaa aca gct gaa aaa ata acc cag cag cga ctg aac ttc tac agg 3308
 Gly Lys Thr Ala Glu Lys Ile Thr Gln Gln Arg Leu Asn Leu Tyr Arg
 1015 1020 1025

agg tca gaa aga caa gag ctt gct gaa ctt gtg aaa aat aga atg aaa 3356
 Arg Ser Glu Arg Gln Glu Leu Ala Glu Leu Val Lys Asn Arg Met Lys
 1030 1035 1040

cac ttg ggt ctt tct aca gtg gga tat gat gaa atg aat gat cat caa 3404
 His Leu Gly Leu Ser Thr Val Gly Tyr Asp Glu Met Asn Asp His Gln
 1045 1050 1055 1060

agt acc ctc tcc tac atc ctg att aac cca tct cca gat acc aga ata 3452
 Ser Thr Leu Ser Tyr Ile Leu Ile Asn Pro Ser Pro Asp Thr Arg Ile
 1065 1070 1075

gag ctg aat gat gtt gta tac tta att cga cca gat cca ctg gcc tac 3500
 Glu Leu Asn Asp Val Val Tyr Leu Ile Arg Pro Asp Pro Leu Ala Tyr
 1080 1085 1090

ctt cca aac agt gag ccc agt cga aga aac agc atc tgc aat gtc act 3548
 Leu Pro Asn Ser Glu Pro Ser Arg Arg Asn Ser Ile Cys Asn Val Thr
 1095 1100 1105

ggt caa gat tct cgg gag gaa act caa ctt tga taaaaataaa atgagaaact 3601
 Gly Gln Asp Ser Arg Glu Glu Thr Gln Leu *
 1110 1115

tttttcctac aaagaccttg cttgaaacca caaaagtgtt gctggcacga aagaaaactag 3661
 atggaaatat atgtaattct ctcataattta aaaacgtaat ctcttctctt agaagtatag 3721
 atcattttga aacttaattgt actacttact ggtactctcc ctattaatat ttgaaggacc 3781
 tcaatggaat aaatttgaaa agctaaatta aaatacaaaa atttaaatct gacatttaat 3841
 tgttttataa taatccaaac tctatgaaag caattttaaa aattattaag gttttatgaa 3901
 gttgacaaaa tctaactata tttggtgcat cacaatggac acagaatgct gctgctcctc 3961
 ttaaaaaatta aatgtgtcat attatattct ttaaacttac tgttttacaa aattgagctc 4021
 atcgtaaattg tctagtcttc tcacatagag attaaccaac aaacttggtg ggctgacttt 4081
 tgtgtaagaa tcatagtttg ctttagaata caaatcttta agtcatttta actttttttt 4141

```

ctgccttacg atataaaaat atttatctta gaatttgaga tggcatagc atgttttatt 4201
acattgaaga aactaaaaca taaatgaaaa gaaacactag gttcctgcac tttttggtaa 4261
ctttatgtct agcaaattatt ttatgccaaag aaaagcatat tataaagcaa atatctatta 4321
ttctcctaaa cgaatgccta gcatagagaa aatacttaat acacatttgt tgacttaaat 4381
ttaattcaag gattgaaaaa ttaactggat atcttgaaat atacagtaat gattgtcctt 4441
agactcttga actttaccat ctttcctatt catatatcta tatagtaaat ttcactagaa 4501
aaattctttt aaaattgaca gaagataatt tatacctttt atggactctg aagacacttc 4561
aaaacattaa aagtccttat gtctttggta atgaaacata cactcaatga ngatgtatta 4621
aattttgact t 4632

```

<210> 8

<211> 1118

<212> PRT

<213> Homo sapiens

<400> 8

```

Met Val Asp Leu Glu Ser Glu Val Pro Pro Leu Pro Pro Arg Tyr Arg
1 5 10 15
Phe Arg Asp Leu Leu Leu Gly Asp Gln Gly Trp Gln Asn Asp Asp Arg
20 25 30
Val Gln Val Glu Phe Tyr Met Asn Glu Asn Thr Phe Lys Glu Arg Leu
35 40 45
Lys Leu Phe Phe Ile Lys Asn Gln Arg Ser Ser Leu Arg Ile Arg Leu
50 55 60
Phe Asn Phe Ser Leu Lys Leu Leu Ser Cys Leu Leu Tyr Ile Ile Arg
65 70 75 80
Val Leu Leu Glu Asn Pro Ser Gln Gly Asn Glu Trp Ser His Ile Phe
85 90 95
Trp Val Asn Arg Ser Leu Pro Leu Trp Gly Leu Gln Val Ser Val Ala
100 105 110
Leu Ile Ser Leu Phe Glu Thr Ile Leu Leu Gly Tyr Leu Ser Tyr Lys
115 120 125
Gly Asn Ile Trp Glu Gln Ile Leu Arg Ile Pro Phe Ile Leu Glu Ile
130 135 140
Ile Asn Ala Val Pro Phe Ile Ile Ser Ile Phe Trp Pro Ser Leu Arg
145 150 155 160
Asn Leu Phe Val Pro Val Phe Leu Asn Cys Trp Leu Ala Lys His Ala
165 170 175
Leu Glu Asn Met Ile Asn Asp Leu His Arg Ala Ile Gln Arg Thr Gln
180 185 190
Ser Ala Met Phe Asn Gln Val Leu Ile Leu Ile Ser Thr Leu Leu Cys
195 200 205
Leu Ile Phe Thr Cys Ile Cys Gly Ile Gln His Leu Glu Arg Ile Gly
210 215 220
Lys Arg Leu Asn Leu Phe Asp Ser Leu Tyr Phe Cys Ile Val Thr Phe
225 230 235 240
Ser Thr Val Gly Phe Gly Asp Val Thr Pro Glu Thr Trp Ser Ser Lys
245 250 255
Leu Phe Val Val Ala Met Ile Cys Val Ala Leu Val Val Leu Pro Ile
260 265 270
Gln Phe Glu Gln Leu Ala Tyr Leu Trp Met Glu Arg Gln Lys Ser Gly
275 280 285
Gly Asn Tyr Ser Arg His Arg Ala Gln Thr Glu Lys His Val Val Leu
290 295 300
Cys Val Ser Ser Leu Lys Ile Asp Leu Leu Met Asp Phe Leu Asn Glu
305 310 315 320
Phe Tyr Ala His Pro Arg Leu Gln Asp Tyr Tyr Val Val Ile Leu Cys
325 330 335
Pro Thr Glu Met Asp Val Gln Val Arg Arg Val Leu Gln Ile Pro Met
340 345 350

```


Trp	Ser	Gln	Arg	Val	Ile	Tyr	Leu	Gln	Gly	Ser	Ala	Leu	Lys	Asp	Gln
		355					360					365			
Asp	Leu	Leu	Arg	Ala	Lys	Met	Asp	Asp	Ala	Glu	Ala	Cys	Phe	Ile	Leu
	370					375					380				
Ser	Ser	Arg	Cys	Glu	Val	Asp	Arg	Thr	Ser	Ser	Asp	His	Gln	Thr	Ile
385					390					395					400
Leu	Arg	Ala	Trp	Ala	Val	Lys	Asp	Phe	Ala	Pro	Asn	Cys	Pro	Leu	Tyr
				405					410					415	
Val	Gln	Ile	Leu	Lys	Pro	Glu	Asn	Lys	Phe	His	Ile	Lys	Phe	Ala	Asp
			420					425					430		
His	Val	Val	Cys	Glu	Glu	Glu	Phe	Lys	Tyr	Ala	Met	Leu	Ala	Leu	Asn
			435				440					445			
Cys	Ile	Cys	Pro	Ala	Thr	Ser	Thr	Leu	Ile	Thr	Leu	Leu	Val	His	Thr
	450					455					460				
Ser	Arg	Gly	Gln	Glu	Gly	Gln	Gln	Ser	Pro	Glu	Gln	Trp	Gln	Lys	Met
465					470					475					480
Tyr	Gly	Arg	Cys	Ser	Gly	Asn	Glu	Val	Tyr	His	Ile	Val	Leu	Glu	Glu
				485					490					495	
Ser	Thr	Phe	Phe	Ala	Glu	Tyr	Glu	Gly	Lys	Ser	Phe	Thr	Tyr	Ala	Ser
			500					505					510		
Phe	His	Ala	His	Lys	Lys	Phe	Gly	Val	Cys	Leu	Ile	Gly	Val	Arg	Arg
		515					520					525			
Glu	Asp	Asn	Lys	Asn	Ile	Leu	Leu	Asn	Pro	Gly	Pro	Arg	Tyr	Ile	Met
	530					535					540				
Asn	Ser	Thr	Asp	Ile	Cys	Phe	Tyr	Ile	Asn	Ile	Thr	Lys	Glu	Glu	Asn
545					550					555					560
Ser	Ala	Phe	Lys	Asn	Gln	Asp	Gln	Gln	Arg	Lys	Ser	Asn	Val	Ser	Arg
				565					570					575	
Ser	Phe	Tyr	His	Gly	Pro	Ser	Arg	Leu	Pro	Val	His	Ser	Ile	Ile	Ala
			580					585					590		
Ser	Met	Gly	Thr	Val	Ala	Ile	Asp	Leu	Gln	Asp	Thr	Ser	Cys	Arg	Ser
		595					600					605			
Ala	Ser	Gly	Pro	Thr	Leu	Ser	Leu	Pro	Thr	Glu	Gly	Ser	Lys	Glu	Ile
	610					615					620				
Arg	Arg	Pro	Ser	Ile	Ala	Pro	Val	Leu	Glu	Val	Ala	Asp	Thr	Ser	Ser
625					630					635					640
Ile	Gln	Thr	Cys	Asp	Leu	Leu	Ser	Asp	Gln	Ser	Glu	Asp	Glu	Thr	Thr
				645					650					655	
Pro	Asp	Glu	Glu	Met	Ser	Ser	Asn	Leu	Glu	Tyr	Ala	Lys	Gly	Tyr	Pro
			660					665					670		
Pro	Tyr	Ser	Pro	Tyr	Ile	Gly	Ser	Ser	Pro	Thr	Phe	Cys	His	Leu	Leu
		675					680					685			
His	Glu	Lys	Val	Pro	Phe	Cys	Cys	Leu	Arg	Leu	Asp	Lys	Ser	Cys	Gln
	690					695					700				
His	Asn	Tyr	Tyr	Glu	Asp	Ala	Lys	Ala	Tyr	Gly	Phe	Lys	Asn	Lys	Leu
705					710					715					720
Ile	Ile	Val	Ala	Ala	Glu	Thr	Ala	Gly	Asn	Gly	Leu	Tyr	Asn	Phe	Ile
				725					730					735	
Val	Pro	Leu	Arg	Ala	Tyr	Tyr	Arg	Pro	Lys	Lys	Glu	Leu	Asn	Pro	Ile
			740					745					750		
Val	Leu	Leu	Leu	Asp	Asn	Pro	Leu	Asp	Asp	Leu	Leu	Arg	Cys	Gly	Val
		755					760					765			
Thr	Phe	Ala	Ala	Asn	Met	Val	Val	Val	Asp	Lys	Glu	Ser	Thr	Met	Ser
					770		775				780				
Ala	Glu	Glu	Asp	Tyr	Met	Ala	Asp	Ala	Lys	Thr	Ile	Val	Asn	Val	Gln
785					790					795					800
Thr	Leu	Phe	Arg	Leu	Phe	Ser	Ser	Leu	Ser	Ile	Ile	Thr	Glu	Leu	Thr
				805					810					815	
His	Pro	Ala	Asn	Met	Arg	Phe	Met	Gln	Phe	Arg	Ala	Lys	Asp	Cys	Tyr
			820					825					830		

```

Ser Leu Ala Leu Ser Lys Leu Glu Lys Lys Glu Arg Glu Arg Gly Ser
835      840      845
Asn Leu Ala Phe Met Phe Arg Leu Pro Phe Ala Ala Gly Arg Val Phe
850      855      860
Ser Ile Ser Met Leu Asp Thr Leu Leu Tyr Gln Ser Phe Val Lys Asp
865      870      875      880
Tyr Met Ile Ser Ile Thr Arg Leu Leu Leu Gly Leu Asp Thr Thr Pro
885      890      895
Gly Ser Gly Phe Leu Cys Ser Met Lys Ile Thr Ala Asp Asp Leu Trp
900      905      910
Ile Arg Thr Tyr Ala Arg Leu Tyr Gln Lys Leu Cys Ser Ser Thr Gly
915      920      925
Asp Val Pro Ile Gly Ile Tyr Arg Thr Glu Ser Gln Lys Leu Thr Thr
930      935      940
Ser Glu Ser Arg Lys Ile Ala Ser Gln Ser Gln Ile Ser Ile Ser Val
945      950      955      960
Glu Glu Trp Glu Asp Thr Lys Asp Ser Lys Glu Gln Gly His His Arg
965      970      975
Ser Asn His Arg Asn Ser Thr Ser Ser Asp Gln Ser Asp His Pro Leu
980      985      990
Leu Arg Arg Lys Ser Met Gln Trp Ala Arg Arg Leu Ser Arg Lys Gly
995      1000      1005
Pro Lys His Ser Gly Lys Thr Ala Glu Lys Ile Thr Gln Gln Arg Leu
1010      1015      1020
Asn Leu Tyr Arg Arg Ser Glu Arg Gln Glu Leu Ala Glu Leu Val Lys
1025      1030      1035      1040
Asn Arg Met Lys His Leu Gly Leu Ser Thr Val Gly Tyr Asp Glu Met
1045      1050      1055
Asn Asp His Gln Ser Thr Leu Ser Tyr Ile Leu Ile Asn Pro Ser Pro
1060      1065      1070
Asp Thr Arg Ile Glu Leu Asn Asp Val Val Tyr Leu Ile Arg Pro Asp
1075      1080      1085
Pro Leu Ala Tyr Leu Pro Asn Ser Glu Pro Ser Arg Arg Asn Ser Ile
1090      1095      1100
Cys Asn Val Thr Gly Gln Asp Ser Arg Glu Glu Thr Gln Leu
1105      1110      1115

```

<210> 9

<211> 3357

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(3357)

<400> 9

```

atg gtt gat ttg gag agc gaa gtg ccc cct ctg cct ccc agg tac agg 48
Met Val Asp Leu Glu Ser Glu Val Pro Pro Leu Pro Pro Arg Tyr Arg
1      5      10      15

```

```

ttt cga gat ttg ctg cta ggg gac caa gga tgg caa aac gat gac aga 96
Phe Arg Asp Leu Leu Leu Gly Asp Gln Gly Trp Gln Asn Asp Asp Arg
20      25      30

```

```

gta caa gtt gaa ttc tat atg aat gaa aat aca ttt aaa gaa aga cta 144
Val Gln Val Glu Phe Tyr Met Asn Glu Asn Thr Phe Lys Glu Arg Leu
35      40      45

```

aaa tta ttt ttc ata aaa aac cag aga tca agt cta agg ata cgc ctg	192
Lys Leu Phe Phe Ile Lys Asn Gln Arg Ser Ser Leu Arg Ile Arg Leu	
50 55 60	
ttc aat ttt tct ctc aaa tta cta agc tgc tta tta tac ata atc cga	240
Phe Asn Phe Ser Leu Lys Leu Leu Ser Cys Leu Leu Tyr Ile Ile Arg	
65 70 75 80	
gta cta cta gaa aac cct tca caa gga aat gaa tgg tct cat atc ttt	288
Val Leu Leu Glu Asn Pro Ser Gln Gly Asn Glu Trp Ser His Ile Phe	
85 90 95	
tgg gtg aac aga agt cta cct ttg tgg ggc tta cag gtt tca gtg gca	336
Trp Val Asn Arg Ser Leu Pro Leu Trp Gly Leu Gln Val Ser Val Ala	
100 105 110	
ttg ata agt ctg ttt gaa aca ata tta ctt ggt tat ctt agt tat aag	384
Leu Ile Ser Leu Phe Glu Thr Ile Leu Leu Gly Tyr Leu Ser Tyr Lys	
115 120 125	
gga aac atc tgg gaa cag att tta cga ata ccc ttc atc ttg gaa ata	432
Gly Asn Ile Trp Glu Gln Ile Leu Arg Ile Pro Phe Ile Leu Glu Ile	
130 135 140	
att aat gca gtt ccc ttc att atc tca ata ttc tgg cct tcc tta agg	480
Ile Asn Ala Val Pro Phe Ile Ile Ser Ile Phe Trp Pro Ser Leu Arg	
145 150 155 160	
aat cta ttt gtc cca gtc ttt ctg aac tgt tgg ctt gcc aaa cat gcc	528
Asn Leu Phe Val Pro Val Phe Leu Asn Cys Trp Leu Ala Lys His Ala	
165 170 175	
ttg gaa aat atg att aat gat cta cac aga gcc att cag cgt aca cag	576
Leu Glu Asn Met Ile Asn Asp Leu His Arg Ala Ile Gln Arg Thr Gln	
180 185 190	
tct gca atg ttt aat caa gtt ttg att tta ata tct aca tta cta tgc	624
Ser Ala Met Phe Asn Gln Val Leu Ile Leu Ile Ser Thr Leu Leu Cys	
195 200 205	
ctt atc ttc acc tgc att tgt ggg atc caa cat ctg gaa cga ata gga	672
Leu Ile Phe Thr Cys Ile Cys Gly Ile Gln His Leu Glu Arg Ile Gly	
210 215 220	
aag agg ctg aat ctc ttt gac tcc ctt tat ttc tgc att gtg acg ttt	720
Lys Arg Leu Asn Leu Phe Asp Ser Leu Tyr Phe Cys Ile Val Thr Phe	
225 230 235 240	
tct act gtg ggc ttc ggg gat gtc act cct gaa aca tgg tcc tcc aag	768
Ser Thr Val Gly Phe Gly Asp Val Thr Pro Glu Thr Trp Ser Ser Lys	
245 250 255	
ctt ttt gta gtt gct atg att tgt gtt gct ctt gtg gtt cta ccc ata	816
Leu Phe Val Val Ala Met Ile Cys Val Ala Leu Val Val Leu Pro Ile	
260 265 270	
cag ttt gaa cag ctg gct tat ttg tgg atg gag aga caa aag tca gga	864
Gln Phe Glu Gln Leu Ala Tyr Leu Trp Met Glu Arg Gln Lys Ser Gly	
275 280 285	

gga aac tat agt cga cat aga gct caa act gaa aag cat gtc gtc ctg	912
Gly Asn Tyr Ser Arg His Arg Ala Gln Thr Glu Lys His Val Val Leu	
290 295 300	
tgt gtc agc tca ctg aag att gat tta ctt atg gat ttt tta aat gaa	960
Cys Val Ser Ser Leu Lys Ile Asp Leu Leu Met Asp Phe Leu Asn Glu	
305 310 315 320	
ttc tat gct cat cct agg ctc cag gat tat tat gtg gtg att ttg tgt	1008
Phe Tyr Ala His Pro Arg Leu Gln Asp Tyr Tyr Val Val Ile Leu Cys	
325 330 335	
cct act gaa atg gat gta cag gtt cga agg gta ctg cag att cca atg	1056
Pro Thr Glu Met Asp Val Gln Val Arg Arg Val Leu Gln Ile Pro Met	
340 345 350	
tgg tcc caa cga gtt atc tac ctt caa ggt tca gcc ctt aaa gat caa	1104
Trp Ser Gln Arg Val Ile Tyr Leu Gln Gly Ser Ala Leu Lys Asp Gln	
355 360 365	
gac cta ttg aga gca aag atg gat gac gct gag gcc tgt ttt att ctc	1152
Asp Leu Leu Arg Ala Lys Met Asp Asp Ala Glu Ala Cys Phe Ile Leu	
370 375 380	
agt agc cgt tgt gaa gtg gat agg aca tca tct gat cac caa aca att	1200
Ser Ser Arg Cys Glu Val Asp Arg Thr Ser Ser Asp His Gln Thr Ile	
385 390 395 400	
ttg aga gca tgg gct gtg aaa gat ttt gct cca aat tgt cct ttg tat	1248
Leu Arg Ala Trp Ala Val Lys Asp Phe Ala Pro Asn Cys Pro Leu Tyr	
405 410 415	
gtc cag ata tta aag cct gaa aat aaa ttt cac atc aaa ttt gct gat	1296
Val Gln Ile Leu Lys Pro Glu Asn Lys Phe His Ile Lys Phe Ala Asp	
420 425 430	
cat gtt gtt tgt gaa gaa gag ttt aaa tac gcc atg tta gct tta aac	1344
His Val Val Cys Glu Glu Glu Phe Lys Tyr Ala Met Leu Ala Leu Asn	
435 440 445	
tgt ata tgc cca gca aca tct aca ctt att aca cta ctg gtt cat acc	1392
Cys Ile Cys Pro Ala Thr Ser Thr Leu Ile Thr Leu Leu Val His Thr	
450 455 460	
tct aga ggg caa gaa ggc cag caa tog cca gaa caa tgg cag aag atg	1440
Ser Arg Gly Gln Glu Gly Gln Gln Ser Pro Glu Gln Trp Gln Lys Met	
465 470 475 480	
tac ggt aga tgc tcc ggg aat gaa gtc tac cac att gtt ttg gaa gaa	1488
Tyr Gly Arg Cys Ser Gly Asn Glu Val Tyr His Ile Val Leu Glu Glu	
485 490 495	
agt aca ttt ttt gct gaa tat gaa gga aag agt ttt aca tat gcc tct	1536
Ser Thr Phe Phe Ala Glu Tyr Glu Gly Lys Ser Phe Thr Tyr Ala Ser	
500 505 510	
ttc cat gca cac aaa aag ttt ggc gtc tgc ttg att ggt gtt agg agg	1584
Phe His Ala His Lys Lys Phe Gly Val Cys Leu Ile Gly Val Arg Arg	
515 520 525	

gag gat aat aaa aac att ttg ctg aat cca ggt cct cga tac att atg	1632
Glu Asp Asn Lys Asn Ile Leu Leu Asn Pro Gly Pro Arg Tyr Ile Met	
530 535 540	
aat tct acg gac ata tgc ttt tat att aat att acc aaa gaa gag aat	1680
Asn Ser Thr Asp Ile Cys Phe Tyr Ile Asn Ile Thr Lys Glu Glu Asn	
545 550 555 560	
tca gca ttt aaa aac caa gac cag cag aga aaa agc aat gtg tcc agg	1728
Ser Ala Phe Lys Asn Gln Asp Gln Gln Arg Lys Ser Asn Val Ser Arg	
565 570 575	
tcg ttt tat cat gga cct tcc aga tta cct gta cat agc ata att gcc	1776
Ser Phe Tyr His Gly Pro Ser Arg Leu Pro Val His Ser Ile Ile Ala	
580 585 590	
agc atg ggt act gtg gct ata gac ctg caa gat aca agc tgt aga tca	1824
Ser Met Gly Thr Val Ala Ile Asp Leu Gln Asp Thr Ser Cys Arg Ser	
595 600 605	
gca agt ggc cct acc ctg tct ctt cct aca gag gga agc aaa gaa ata	1872
Ala Ser Gly Pro Thr Leu Ser Leu Pro Thr Glu Gly Ser Lys Glu Ile	
610 615 620	
aga aga cct agc att gct cct gtt tta gag gtt gca gat aca tca tcg	1920
Arg Arg Pro Ser Ile Ala Pro Val Leu Glu Val Ala Asp Thr Ser Ser	
625 630 635 640	
att caa aca tgt gat ctt cta agt gac caa tca gaa gat gaa act aca	1968
Ile Gln Thr Cys Asp Leu Leu Ser Asp Gln Ser Glu Asp Glu Thr Thr	
645 650 655	
cca gat gaa gaa atg tct tca aac tta gag tat gct aaa ggt tac cca	2016
Pro Asp Glu Glu Met Ser Ser Asn Leu Glu Tyr Ala Lys Gly Tyr Pro	
660 665 670	
cct tat tct cca tat ata gga agt tca ccc act ttt tgt cat ctc ctt	2064
Pro Tyr Ser Pro Tyr Ile Gly Ser Ser Pro Thr Phe Cys His Leu Leu	
675 680 685	
cat gaa aaa gta cca ttt tgc tgc tta aga tta gac aag agt tgc caa	2112
His Glu Lys Val Pro Phe Cys Cys Leu Arg Leu Asp Lys Ser Cys Gln	
690 695 700	
cat aac tac tat gag gat gca aaa gcc tat gga ttc aaa aat aaa cta	2160
His Asn Tyr Tyr Glu Asp Ala Lys Ala Tyr Gly Phe Lys Asn Lys Leu	
705 710 715 720	
att ata gtt gca gct gaa aca gct gga aat gga tta tat aac ttt att	2208
Ile Ile Val Ala Ala Glu Thr Ala Gly Asn Gly Leu Tyr Asn Phe Ile	
725 730 735	
gtt cct ctc agg gca tat tat aga cca aag aaa gaa ctt aat ccc ata	2256
Val Pro Leu Arg Ala Tyr Tyr Arg Pro Lys Lys Glu Leu Asn Pro Ile	
740 745 750	
gta ctg cta ttg gat aac ccc cta gat gac tta ctc agg tgt gga gtg	2304
Val Leu Leu Leu Asp Asn Pro Leu Asp Asp Leu Leu Arg Cys Gly Val	
755 760 765	

act ttt gct gct aat atg gtg gtt gtg gat aaa gag agc acc atg agt	2352
Thr Phe Ala Ala Asn Met Val Val Val Asp Lys Glu Ser Thr Met Ser	
770 775 780	
gcc gag gaa gac tac atg gca gat gcc aaa acc att gtg aac gtg cag	2400
Ala Glu Glu Asp Tyr Met Ala Asp Ala Lys Thr Ile Val Asn Val Gln	
785 790 795 800	
aca ctc ttc agg ttg ttt tcc agt ctc agt att atc aca gag cta act	2448
Thr Leu Phe Arg Leu Phe Ser Ser Leu Ser Ile Ile Thr Glu Leu Thr	
805 810 815	
cac ccc gcc aac atg aga ttc atg caa ttc aga gcc aaa gac tgt tac	2496
His Pro Ala Asn Met Arg Phe Met Gln Phe Arg Ala Lys Asp Cys Tyr	
820 825 830	
tct ctt gct ctt tca aaa ctg gaa aag aaa gaa cgg gag aga ggc tct	2544
Ser Leu Ala Leu Ser Lys Leu Glu Lys Lys Glu Arg Glu Arg Gly Ser	
835 840 845	
aac ttg gcc ttt atg ttt cga ctg cct ttt gct gct ggg agg gtg ttt	2592
Asn Leu Ala Phe Met Phe Arg Leu Pro Phe Ala Ala Gly Arg Val Phe	
850 855 860	
agc atc agt atg ttg gac act ctg ctg tat cag tca ttt gtg aag gat	2640
Ser Ile Ser Met Leu Asp Thr Leu Leu Tyr Gln Ser Phe Val Lys Asp	
865 870 875 880	
tat atg att tct atc acg aga ctt ctg ttg gga ctg gac act aca cca	2688
Tyr Met Ile Ser Ile Thr Arg Leu Leu Leu Gly Leu Asp Thr Thr Pro	
885 890 895	
gga tct ggg ttt ctt tgt tct atg aaa atc act gca gat gac tta tgg	2736
Gly Ser Gly Phe Leu Cys Ser Met Lys Ile Thr Ala Asp Asp Leu Trp	
900 905 910	
atc aga act tat gcc aga ctt tat cag aag ttg tgt tct tct act gga	2784
Ile Arg Thr Tyr Ala Arg Leu Tyr Gln Lys Leu Cys Ser Ser Thr Gly	
915 920 925	
gat gtt ccc att gga atc tac agg act gag tct cag aaa ctt act aca	2832
Asp Val Pro Ile Gly Ile Tyr Arg Thr Glu Ser Gln Lys Leu Thr Thr	
930 935 940	
tct gag tct cga aaa ata gca tca caa tct caa ata tct atc agt gta	2880
Ser Glu Ser Arg Lys Ile Ala Ser Gln Ser Gln Ile Ser Ile Ser Val	
945 950 955 960	
gaa gag tgg gaa gac acc aaa gac tcc aaa gaa caa ggg cac cac cgc	2928
Glu Glu Trp Glu Asp Thr Lys Asp Ser Lys Glu Gln Gly His His Arg	
965 970 975	
agc aac cac cgc aac tca aca tcc agt gac cag tcg gac cat ccc ttg	2976
Ser Asn His Arg Asn Ser Thr Ser Ser Asp Gln Ser Asp His Pro Leu	
980 985 990	
ctg cgg aga aaa agc atg cag tgg gcc cga aga ctg agc aga aaa ggc	3024
Leu Arg Arg Lys Ser Met Gln Trp Ala Arg Arg Leu Ser Arg Lys Gly	
995 1000 1005	

cca aaa cac tct ggt aaa aca gct gaa aaa ata acc cag cag cga ctg 3072
 Pro Lys His Ser Gly Lys Thr Ala Glu Lys Ile Thr Gln Gln Arg Leu
 1010 1015 1020

aac ctc tac agg agg tca gaa aga caa gag ctt gct gaa ctt gtg aaa 3120
 Asn Leu Tyr Arg Arg Ser Glu Arg Gln Glu Leu Ala Glu Leu Val Lys
 1025 1030 1035 1040

aat aga atg aaa cac ttg ggt ctt tct aca gtg gga tat gat gaa atg 3168
 Asn Arg Met Lys His Leu Gly Leu Ser Thr Val Gly Tyr Asp Glu Met
 1045 1050 1055

aat gat cat caa agt acc ctc tcc tac atc ctg att aac cca tct cca 3216
 Asn Asp His Gln Ser Thr Leu Ser Tyr Ile Leu Ile Asn Pro Ser Pro
 1060 1065 1070

gat acc aga ata gag ctg aat gat gtt gta tac tta att cga cca gat 3264
 Asp Thr Arg Ile Glu Leu Asn Asp Val Val Tyr Leu Ile Arg Pro Asp
 1075 1080 1085

cca ctg gcc tac ctt cca aac agt gag ccc agt cga aga aac agc atc 3312
 Pro Leu Ala Tyr Leu Pro Asn Ser Glu Pro Ser Arg Arg Asn Ser Ile
 1090 1095 1100

tgc aat gtc act ggt caa gat tct cgg gag gaa act caa ctt tga 3357
 Cys Asn Val Thr Gly Gln Asp Ser Arg Glu Glu Thr Gln Leu *
 1105 1110 1115

<210> 10

<211> 2847

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (561)...(2477)

<400> 10

ccacgcgtcc ggccctgtgc ttccggtggc ggccggaggc tgatggcgag tgggtgctgaa 60
 gggacagctc cagcagtggc tgatttgagg gagaaacaaa atctgcagat ggaatccgag 120
 caggcgact tcaccttcaa gtggtgagct ctctgacct gcggccagtc tccactccat 180
 tcacggccag ccgatctgcc cgctcccga ggggtcggc agtgccggct ggacccgccc 240
 cgagctccat ggtttgccc accctgcgcg atggtgactc tgggcgcgga gggtggcgac 300
 tggcaaattc gcagatcaca gaatgaaggc ggggagcgcg gccggcgccc ggccggggct 360
 ttctcccca cccagcgcg cagggaagcg gctcaaccac ctgaatccgg aaaacgccaa 420
 caagtagttt ctgctcggag aaggcgcgct cacctgggcg ccaagactca gtcccgtgc 480
 ccagagaacc tcgtccactc ggaaacccaa gcagaaccac tttctctctg gtctcgtaa 540
 gtcattgtctg agtcacagag atg ggc aag atc gag aac aac gag agg gtg atc 593
 Met Gly Lys Ile Glu Asn Asn Glu Arg Val Ile
 1 5 10

ctc aat gtc ggg ggc acc cgg cac gaa acc tac cgc agc acc ctc aag 641
 Leu Asn Val Gly Gly Thr Arg His Glu Thr Tyr Arg Ser Thr Leu Lys
 15 20 25

acc ctg cct gga aca cgc ctg gcc ctt ctt gcc tcc tcc gag ccc cca 689
 Thr Leu Pro Gly Thr Arg Leu Ala Leu Leu Ala Ser Ser Glu Pro Pro
 30 35 40

ggc gac tgc ttg acc acg gcg ggc gac aag ctg cag ccg tgc ccg cct	737
Gly Asp Cys Leu Thr Thr Ala Gly Asp Lys Leu Gln Pro Ser Pro Pro	
45 50 55	
cca ctg tgc ccg ccg ccg aga gcg ccc ccg ctg tcc ccc ggg cca ggc	785
Pro Leu Ser Pro Pro Pro Arg Ala Pro Pro Leu Ser Pro Gly Pro Gly	
60 65 70 75	
ggc tgc ttc gag ggc ggc gcg ggc aac tgc agt tcc cgc ggc ggc agg	833
Gly Cys Phe Glu Gly Gly Ala Gly Asn Cys Ser Ser Arg Gly Gly Arg	
80 85 90	
gcc agc gac cat ccc ggt ggc ggc cgc gag ttc ttc ttc gac cgg cac	881
Ala Ser Asp His Pro Gly Gly Gly Arg Glu Phe Phe Phe Asp Arg His	
95 100 105	
ccg ggc gtc ttc gcc tat gtg ctc aat tac tac cgc acc ggc aag ctg	929
Pro Gly Val Phe Ala Tyr Val Leu Asn Tyr Tyr Arg Thr Gly Lys Leu	
110 115 120	
cac tgc ccc gca gac gtg tgc ggc ccg ctc ttc gag gag gag ctg gcc	977
His Cys Pro Ala Asp Val Cys Gly Pro Leu Phe Glu Glu Glu Leu Ala	
125 130 135	
ttc tgg ggc atc gac gag acc gac gtg gag ccc tgc tgc tgg atg acc	1025
Phe Trp Gly Ile Asp Glu Thr Asp Val Glu Pro Cys Cys Trp Met Thr	
140 145 150 155	
tac cgg cag cac cgc gac gcc gag gag gcg ctg gac atc ttc gag acc	1073
Tyr Arg Gln His Arg Asp Ala Glu Glu Ala Leu Asp Ile Phe Glu Thr	
160 165 170	
ccc gac ctc att ggc ggc gac ccc ggc gac gac gag gac ctg gcg gcc	1121
Pro Asp Leu Ile Gly Gly Asp Pro Gly Asp Asp Glu Asp Leu Ala Ala	
175 180 185	
aag agg ctg ggc atc gag gac gcg gcg ggc ctc ggc ggc ccc gac ggc	1169
Lys Arg Leu Gly Ile Glu Asp Ala Ala Gly Leu Gly Gly Pro Asp Gly	
190 195 200	
aaa tct ggc cgc tgg agg agg ctg cag ccc cgc atg tgg gcc ctc ttc	1217
Lys Ser Gly Arg Trp Arg Arg Leu Gln Pro Arg Met Trp Ala Leu Phe	
205 210 215	
gaa gac ccc tac tgc tcc aga gcc gcc agg ttt att gct ttt gct tct	1265
Glu Asp Pro Tyr Ser Ser Arg Ala Ala Arg Phe Ile Ala Phe Ala Ser	
220 225 230 235	
tta ttc ttc atc ctg gtt tca att aca act ttt tgc ctg gaa aca cat	1313
Leu Phe Phe Ile Leu Val Ser Ile Thr Thr Phe Cys Leu Glu Thr His	
240 245 250	
gaa gct ttc aat att gtt aaa aac aag aca gaa cca gtc atc aat ggc	1361
Glu Ala Phe Asn Ile Val Lys Asn Lys Thr Glu Pro Val Ile Asn Gly	
255 260 265	
aca agt gtt gtt cta cag tat gaa att gaa acg gat cct gcc ttg acg	1409
Thr Ser Val Val Leu Gln Tyr Glu Ile Glu Thr Asp Pro Ala Leu Thr	
270 275 280	

tat gta gaa gga gtg tgt gtg gtg tgg ttt act ttt gaa ttt tta gtc	1457
Tyr Val Glu Gly Val Cys Val Val Trp Phe Thr Phe Glu Phe Leu Val	
285 290 295	
cgt att gtt ttt tca ccc aat aaa ctt gaa ttc atc aaa aat ctc ttg	1505
Arg Ile Val Phe Ser Pro Asn Lys Leu Glu Phe Ile Lys Asn Leu Leu	
300 305 310 315	
aat atc att gac ttt gtg gcc atc cta cct ttc tac tta gag gtg gga	1553
Asn Ile Ile Asp Phe Val Ala Ile Leu Pro Phe Tyr Leu Glu Val Gly	
320 325 330	
ctc agt ggg ctg tca tcc aaa gct gct aaa gat gtg ctt ggc ttc ctc	1601
Leu Ser Gly Leu Ser Ser Lys Ala Ala Lys Asp Val Leu Gly Phe Leu	
335 340 345	
agg gtg gta agg ttt gtg agg atc ctg aga att ttc aag ctc acc cgc	1649
Arg Val Val Arg Phe Val Arg Ile Leu Arg Ile Phe Lys Leu Thr Arg	
350 355 360	
cat ttt gta ggt ctg agg gtg ctt gga cat act ctt cga gct agt act	1697
His Phe Val Gly Leu Arg Val Leu Gly His Thr Leu Arg Ala Ser Thr	
365 370 375	
aat gaa ttt ttg ctg ctg ata att ttc ctg gct cta gga gtt ttg ata	1745
Asn Glu Phe Leu Leu Leu Ile Ile Phe Leu Ala Leu Gly Val Leu Ile	
380 385 390 395	
ttt gct acc atg atc tac tat gcc gag aga gtg gga gct caa cct aac	1793
Phe Ala Thr Met Ile Tyr Tyr Ala Glu Arg Val Gly Ala Gln Pro Asn	
400 405 410	
gac cct tca gct agt gag cac aca cag ttc aaa aac att ccc att ggg	1841
Asp Pro Ser Ala Ser Glu His Thr Gln Phe Lys Asn Ile Pro Ile Gly	
415 420 425	
ttc tgg tgg gct gta gtg acc atg act acc ctg ggt tat ggg gat atg	1889
Phe Trp Trp Ala Val Val Thr Met Thr Thr Leu Gly Tyr Gly Asp Met	
430 435 440	
tac ccc caa aca tgg tca gcc atg ctg gtg gga gcc ctg tgt gct ctg	1937
Tyr Pro Gln Thr Trp Ser Gly Met Leu Val Gly Ala Leu Cys Ala Leu	
445 450 455	
gct gga gtg ctg aca ata gcc atg cca gtg cct gtc att gtc aat aat	1985
Ala Gly Val Leu Thr Ile Ala Met Pro Val Pro Val Ile Val Asn Asn	
460 465 470 475	
ttt gga atg tac tac tcc ttg gca atg gca aag cag aaa ctt cca agg	2033
Phe Gly Met Tyr Tyr Ser Leu Ala Met Ala Lys Gln Lys Leu Pro Arg	
480 485 490	
aaa aga aag aag cac atc cct cct gct cct cag gca agc tca cct act	2081
Lys Arg Lys Lys His Ile Pro Pro Ala Pro Gln Ala Ser Ser Pro Thr	
495 500 505	
ttt tgc aag aca gaa tta aat atg gcc tgc aat agt aca cag agt gac	2129
Phe Cys Lys Thr Glu Leu Asn Met Ala Cys Asn Ser Thr Gln Ser Asp	
510 515 520	

aca tgt ctg ggc aaa gac aat cga ctt ctg gaa cat aac aga tca gtg 2177
 Thr Cys Leu Gly Lys Asp Asn Arg Leu Leu Glu His Asn Arg Ser Val
 525 530 535

tta tca ggt gac gac agt aca gga agt gag ccg cca cta tca ccc cca 2225
 Leu Ser Gly Asp Asp Ser Thr Gly Ser Glu Pro Pro Leu Ser Pro Pro
 540 545 550 555

gaa agg ctc ccc atc aga cgc tct agt acc aga gac aaa aac aga aga 2273
 Glu Arg Leu Pro Ile Arg Arg Ser Ser Thr Arg Asp Lys Asn Arg Arg
 560 565 570

ggg gaa aca tgt ttc cta ctg acg aca ggt gat tac acg tgt gct tct 2321
 Gly Glu Thr Cys Phe Leu Leu Thr Thr Gly Asp Tyr Thr Cys Ala Ser
 575 580 585

gat gga ggg atc agg aaa gga tat gaa aaa tcc cga agc tta aac aac 2369
 Asp Gly Gly Ile Arg Lys Gly Tyr Glu Lys Ser Arg Ser Leu Asn Asn
 590 595 600

ata gcg ggc ttg gca ggc aat gct ctg agg ctc tct cca gta aca tca 2417
 Ile Ala Gly Leu Ala Gly Asn Ala Leu Arg Leu Ser Pro Val Thr Ser
 605 610 615

ccc tac aac tct cct tgt cct ctg agg cgc tct cga tct ccc atc cca 2465
 Pro Tyr Asn Ser Pro Cys Pro Leu Arg Arg Ser Arg Ser Pro Ile Pro
 620 625 630 635

tct atc ttg taa accaaacaac caaactgcat cagtcggcta aattgtatta 2517
 Ser Ile Leu *

attcaagygc tgtttaccce ataattggaaa taattaaatg tagagttact ccaggtccca 2577
 ttaatacagt ataaatcttg cgtgatacta caatttgaag tcagaaatgc cacttgggta 2637
 gctaataaat cttaccagg ctttaaagat tgtctaaagt agtgctaaga tccctcctat 2697
 taattgccct gatatccttt tgcaataaaa tgacagatag tgtcagatat tgaccagtgc 2757
 actaatatat aaacataccc tcagggagat atatttaaaa cagtgtgctt ccaaattgcca 2817
 accacttcat tggaacttta tttcttgtga 2847

<210> 11

<211> 638

<212> PRT

<213> Homo sapiens

<400> 11

Met Gly Lys Ile Glu Asn Asn Glu Arg Val Ile Leu Asn Val Gly Gly
 1 5 10 15
 Thr Arg His Glu Thr Tyr Arg Ser Thr Leu Lys Thr Leu Pro Gly Thr
 20 25 30
 Arg Leu Ala Leu Leu Ala Ser Ser Glu Pro Pro Gly Asp Cys Leu Thr
 35 40 45
 Thr Ala Gly Asp Lys Leu Gln Pro Ser Pro Pro Pro Leu Ser Pro Pro
 50 55 60
 Pro Arg Ala Pro Pro Leu Ser Pro Gly Pro Gly Gly Cys Phe Glu Gly
 65 70 75 80
 Gly Ala Gly Asn Cys Ser Ser Arg Gly Gly Arg Ala Ser Asp His Pro
 85 90 95
 Gly Gly Gly Arg Glu Phe Phe Phe Asp Arg His Pro Gly Val Phe Ala
 100 105 110
 Tyr Val Leu Asn Tyr Tyr Arg Thr Gly Lys Leu His Cys Pro Ala Asp
 115 120 125

Val	Cys	Gly	Pro	Leu	Phe	Glu	Glu	Glu	Leu	Ala	Phe	Trp	Gly	Ile	Asp
130						135					140				
Glu	Thr	Asp	Val	Glu	Pro	Cys	Cys	Trp	Met	Thr	Tyr	Arg	Gln	His	Arg
145					150					155					160
Asp	Ala	Glu	Glu	Ala	Leu	Asp	Ile	Phe	Glu	Thr	Pro	Asp	Leu	Ile	Gly
				165					170					175	
Gly	Asp	Pro	Gly	Asp	Asp	Glu	Asp	Leu	Ala	Ala	Lys	Arg	Leu	Gly	Ile
			180					185					190		
Glu	Asp	Ala	Ala	Gly	Leu	Gly	Gly	Pro	Asp	Gly	Lys	Ser	Gly	Arg	Trp
		195					200					205			
Arg	Arg	Leu	Gln	Pro	Arg	Met	Trp	Ala	Leu	Phe	Glu	Asp	Pro	Tyr	Ser
	210					215					220				
Ser	Arg	Ala	Ala	Arg	Phe	Ile	Ala	Phe	Ala	Ser	Leu	Phe	Phe	Ile	Leu
225					230					235					240
Val	Ser	Ile	Thr	Thr	Phe	Cys	Leu	Glu	Thr	His	Glu	Ala	Phe	Asn	Ile
			245						250					255	
Val	Lys	Asn	Lys	Thr	Glu	Pro	Val	Ile	Asn	Gly	Thr	Ser	Val	Val	Leu
		260						265					270		
Gln	Tyr	Glu	Ile	Glu	Thr	Asp	Pro	Ala	Leu	Thr	Tyr	Val	Glu	Gly	Val
	275						280					285			
Cys	Val	Val	Trp	Phe	Thr	Phe	Glu	Phe	Leu	Val	Arg	Ile	Val	Phe	Ser
	290					295					300				
Pro	Asn	Lys	Leu	Glu	Phe	Ile	Lys	Asn	Leu	Leu	Asn	Ile	Ile	Asp	Phe
305					310					315					320
Val	Ala	Ile	Leu	Pro	Phe	Tyr	Leu	Glu	Val	Gly	Leu	Ser	Gly	Leu	Ser
			325						330					335	
Ser	Lys	Ala	Ala	Lys	Asp	Val	Leu	Gly	Phe	Leu	Arg	Val	Val	Arg	Phe
		340						345					350		
Val	Arg	Ile	Leu	Arg	Ile	Phe	Lys	Leu	Thr	Arg	His	Phe	Val	Gly	Leu
	355					360						365			
Arg	Val	Leu	Gly	His	Thr	Leu	Arg	Ala	Ser	Thr	Asn	Glu	Phe	Leu	Leu
	370					375					380				
Leu	Ile	Ile	Phe	Leu	Ala	Leu	Gly	Val	Leu	Ile	Phe	Ala	Thr	Met	Ile
385					390					395					400
Tyr	Tyr	Ala	Glu	Arg	Val	Gly	Ala	Gln	Pro	Asn	Asp	Pro	Ser	Ala	Ser
			405						410					415	
Glu	His	Thr	Gln	Phe	Lys	Asn	Ile	Pro	Ile	Gly	Phe	Trp	Trp	Ala	Val
		420						425					430		
Val	Thr	Met	Thr	Thr	Leu	Gly	Tyr	Gly	Asp	Met	Tyr	Pro	Gln	Thr	Trp
	435					440						445			
Ser	Gly	Met	Leu	Val	Gly	Ala	Leu	Cys	Ala	Leu	Ala	Gly	Val	Leu	Thr
	450					455					460				
Ile	Ala	Met	Pro	Val	Pro	Val	Ile	Val	Asn	Asn	Phe	Gly	Met	Tyr	Tyr
465					470					475					480
Ser	Leu	Ala	Met	Ala	Lys	Gln	Lys	Leu	Pro	Arg	Lys	Arg	Lys	Lys	His
			485						490					495	
Ile	Pro	Pro	Ala	Pro	Gln	Ala	Ser	Ser	Pro	Thr	Phe	Cys	Lys	Thr	Glu
		500						505					510		
Leu	Asn	Met	Ala	Cys	Asn	Ser	Thr	Gln	Ser	Asp	Thr	Cys	Leu	Gly	Lys
	515						520					525			
Asp	Asn	Arg	Leu	Leu	Glu	His	Asn	Arg	Ser	Val	Leu	Ser	Gly	Asp	Asp
	530					535					540				
Ser	Thr	Gly	Ser	Glu	Pro	Pro	Leu	Ser	Pro	Pro	Glu	Arg	Leu	Pro	Ile
545					550					555					560
Arg	Arg	Ser	Ser	Thr	Arg	Asp	Lys	Asn	Arg	Arg	Gly	Glu	Thr	Cys	Phe
			565						570					575	
Leu	Leu	Thr	Thr	Gly	Asp	Tyr	Thr	Cys	Ala	Ser	Asp	Gly	Gly	Ile	Arg
		580						585					590		
Lys	Gly	Tyr	Glu	Lys	Ser	Arg	Ser	Leu	Asn	Asn	Ile	Ala	Gly	Leu	Ala
	595						600					605			

Gly Asn Ala Leu Arg Leu Ser Pro Val Thr Ser Pro Tyr Asn Ser Pro
 610 615 620
 Cys Pro Leu Arg Arg Ser Arg Ser Pro Ile Pro Ser Ile Leu
 625 630 635

<210> 12

<211> 1917

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)... (1917)

<400> 12

atg ggc aag atc gag aac aac gag agg gtg atc ctc aat gtc ggg ggc	48
Met Gly Lys Ile Glu Asn Asn Glu Arg Val Ile Leu Asn Val Gly Gly	
1 5 10 15	
acc cgg cac gaa acc tac cgc agc acc ctc aag acc ctg cct gga aca	96
Thr Arg His Glu Thr Tyr Arg Ser Thr Leu Lys Thr Leu Pro Gly Thr	
20 25 30	
cgc ctg gcc ctt ctt gcc tcc tcc gag ccc cca ggc gac tgc ttg acc	144
Arg Leu Ala Leu Leu Ala Ser Ser Glu Pro Pro Gly Asp Cys Leu Thr	
35 40 45	
acg gcg ggc gac aag ctg cag ccg tcg ccg cct cca ctg tcg ccg ccg	192
Thr Ala Gly Asp Lys Leu Gln Pro Ser Pro Pro Pro Leu Ser Pro Pro	
50 55 60	
ccg aga gcg ccc ccg ctg tcc ccc ggg cca ggc ggc tgc ttc gag ggc	240
Pro Arg Ala Pro Pro Leu Ser Pro Gly Pro Gly Gly Cys Phe Glu Gly	
65 70 75 80	
ggc gcg ggc aac tgc agt tcc cgc ggc ggc agg gcc agc gac cat ccc	288
Gly Ala Gly Asn Cys Ser Ser Arg Gly Gly Arg Ala Ser Asp His Pro	
85 90 95	
ggt ggc ggc cgc gag ttc ttc ttc gac cgg cac ccg ggc gtc ttc gcc	336
Gly Gly Gly Arg Glu Phe Phe Phe Asp Arg His Pro Gly Val Phe Ala	
100 105 110	
tat gtg ctc aat tac tac cgc acc ggc aag ctg cac tgc ccc gca gac	384
Tyr Val Leu Asn Tyr Tyr Arg Thr Gly Lys Leu His Cys Pro Ala Asp	
115 120 125	
gtg tgc ggg ccg ctc ttc gag gag gag ctg gcc ttc tgg ggc atc gac	432
Val Cys Gly Pro Leu Phe Glu Glu Glu Leu Ala Phe Trp Gly Ile Asp	
130 135 140	
gag acc gac gtg gag ccc tgc tgc tgg atg acc tac cgg cag cac cgc	480
Glu Thr Asp Val Glu Pro Cys Cys Trp Met Thr Tyr Arg Gln His Arg	
145 150 155 160	
gac gcc gag gag gcg ctg gac atc ttc gag acc ccc gac ctc att ggc	528
Asp Ala Glu Glu Ala Leu Asp Ile Phe Glu Thr Pro Asp Leu Ile Gly	
165 170 175	

ggc gac ccc ggc gac gac gag gac ctg gcg gcc aag agg ctg ggc atc Gly Asp Pro Gly Asp Asp Glu Asp Leu Ala Ala Lys Arg Leu Gly Ile 180 185 190	576
gag gac gcg gcg ggg ctc ggg ggc ccc gac ggc aaa tct ggc cgc tgg Glu Asp Ala Ala Gly Leu Gly Gly Pro Asp Gly Lys Ser Gly Arg Trp 195 200 205	624
agg agg ctg cag ccc cgc atg tgg gcc ctc ttc gaa gac ccc tac tcg Arg Arg Leu Gln Pro Arg Met Trp Ala Leu Phe Glu Asp Pro Tyr Ser 210 215 220	672
tcc aga gcc gcc agg ttt att gct ttt gct tct tta ttc ttc atc ctg Ser Arg Ala Ala Arg Phe Ile Ala Phe Ala Ser Leu Phe Phe Ile Leu 225 230 235 240	720
gtt tca att aca act ttt tgc ctg gaa aca cat gaa gct ttc aat att Val Ser Ile Thr Thr Phe Cys Leu Glu Thr His Glu Ala Phe Asn Ile 245 250 255	768
gtt aaa aac aag aca gaa cca gtc atc aat ggc aca agt gtt gtt cta Val Lys Asn Lys Thr Glu Pro Val Ile Asn Gly Thr Ser Val Val Leu 260 265 270	816
cag tat gaa att gaa acg gat cct gcc ttg acg tat gta gaa gga gtg Gln Tyr Glu Ile Glu Thr Asp Pro Ala Leu Thr Tyr Val Glu Gly Val 275 280 285	864
tgt gtg gtg tgg ttt act ttt gaa ttt tta gtc cgt att gtt ttt tca Cys Val Val Trp Phe Thr Phe Glu Phe Leu Val Arg Ile Val Phe Ser 290 295 300	912
ccc aat aaa ctt gaa ttc atc aaa aat ctc ttg aat atc att gac ttt Pro Asn Lys Leu Glu Phe Ile Lys Asn Leu Leu Asn Ile Ile Asp Phe 305 310 315 320	960
gtg gcc atc cta cct ttc tac tta gag gtg gga ctc agt ggg ctg tca Val Ala Ile Leu Pro Phe Tyr Leu Glu Val Gly Leu Ser Gly Leu Ser 325 330 335	1008
tcc aaa gct gct aaa gat gtg ctt ggc ttc ctc agg gtg gta agg ttt Ser Lys Ala Ala Lys Asp Val Leu Gly Phe Leu Arg Val Val Arg Phe 340 345 350	1056
gtg agg atc ctg aga att ttc aag ctc acc cgc cat ttt gta ggt ctg Val Arg Ile Leu Arg Ile Phe Lys Leu Thr Arg His Phe Val Gly Leu 355 360 365	1104
agg gtg ctt gga cat act ctt cga gct agt act aat gaa ttt ttg ctg Arg Val Leu Gly His Thr Leu Arg Ala Ser Thr Asn Glu Phe Leu Leu 370 375 380	1152
ctg ata att ttc ctg gct cta gga gtt ttg ata ttt gct acc atg atc Leu Ile Ile Phe Leu Ala Leu Gly Val Leu Ile Phe Ala Thr Met Ile 385 390 395 400	1200
tac tat gcc gag aga gtg gga gct caa cct aac gac cct tca gct agt Tyr Tyr Ala Glu Arg Val Gly Ala Gln Pro Asn Asp Pro Ser Ala Ser 405 410 415	1248

gag cac aca cag ttc aaa aac att ccc att ggg ttc tgg tgg gct gta	1296
Glú His Thr Gln Phe Lys Asn Ile Pro Ile Gly Phe Trp Trp Ala Val	
420 425 430	
gtg acc atg act acc ctg ggt tat ggg gat atg tac ccc caa aca tgg	1344
Val Thr Met Thr Thr Leu Gly Tyr Gly Asp Met Tyr Pro Gln Thr Trp	
435 440 445	
tca ggc atg ctg gtg gga gcc ctg tgt gct ctg gct gga gtg ctg aca	1392
Ser Gly Met Leu Val Gly Ala Leu Cys Ala Leu Ala Gly Val Leu Thr	
450 455 460	
ata gcc atg cca gtg cct gtc att gtc aat aat ttt gga atg tac tac	1440
Ile Ala Met Pro Val Pro Val Ile Val Asn Asn Phe Gly Met Tyr Tyr	
465 470 475 480	
tcc ttg gca atg gca aag cag aaa ctt cca agg aaa aga aag aag cac	1488
Ser Leu Ala Met Ala Lys Gln Lys Leu Pro Arg Lys Arg Lys Lys His	
485 490 495	
atc cct cct gct cct cag gca agc tca cct act ttt tgc aag aca gaa	1536
Ile Pro Pro Ala Pro Gln Ala Ser Ser Pro Thr Phe Cys Lys Thr Glu	
500 505 510	
tta aat atg gcc tgc aat agt aca cag agt gac aca tgt ctg ggc aaa	1584
Leu Asn Met Ala Cys Asn Ser Thr Gln Ser Asp Thr Cys Leu Gly Lys	
515 520 525	
gac aat cga ctt ctg gaa cat aac aga tca gtg tta tca ggt gac gac	1632
Asp Asn Arg Leu Leu Glu His Asn Arg Ser Val Leu Ser Gly Asp Asp	
530 535 540	
agt aca gga agt gag ccg cca cta tca ccc cca gaa agg ctg ccc atc	1680
Ser Thr Gly Ser Glu Pro Pro Leu Ser Pro Pro Glu Arg Leu Pro Ile	
545 550 555 560	
aga cgc tct agt acc aga gac aaa aac aga aga ggg gaa aca tgt ttc	1728
Arg Arg Ser Ser Thr Arg Asp Lys Asn Arg Arg Gly Glu Thr Cys Phe	
565 570 575	
cta ctg acg aca ggt gat tac acg tgt gct tct gat gga ggg atc agg	1776
Leu Leu Thr Thr Gly Asp Tyr Thr Cys Ala Ser Asp Gly Gly Ile Arg	
580 585 590	
aaa gga tat gaa aaa tcc cga agc tta aac aac ata gcg ggc ttg gca	1824
Lys Gly Tyr Glu Lys Ser Arg Ser Leu Asn Asn Ile Ala Gly Leu Ala	
595 600 605	
ggc aat gct ctg agg ctg tct cca gta aca tca ccc tac aac tct cct	1872
Gly Asn Ala Leu Arg Leu Ser Pro Val Thr Ser Pro Tyr Asn Ser Pro	
610 615 620	
tgt cct ctg agg cgc tct cga tct ccc atc cca tct atc ttg taa	1917
Cys Pro Leu Arg Arg Ser Arg Ser Pro Ile Pro Ser Ile Leu *	
625 630 635	

<210> 13

<211> 6582

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (524)...(3913)

<400> 13

ccacgcgtcc gccacgcgt ccgcccacgc gtccgagaag gcttaggtgg gcaggcagga 60
 cgagagaaag actgagagga gggaaagccg cgtaggtggg agtacagcgg cgcgagggtc 120
 gagggggaac cctcgtcggt gcagatgagg aggggtgggt ttcagaacta gtccccctc 180
 gcaccccgcc ccgccccctc cgcgctgggg tcttcacggt gccctgcctc agagcccgcc 240
 tccaccacgc ccggaagagg gagtctggcc gtccgctggc tcagggcggg ccggttggt 300
 gtaccacaggc tccctggccc gagtgcggga ccagagcgcg gggcggcgcg gcagccgcgg 360
 gccgaggagg ggctgcgagc gaaacggcgc ggcgcgccac ggcgagcagc ttagggccgg 420
 ggcgaggagg gctgtggctc ccgacagaga caggggagta gtgtcgggt gaggcgagac 480
 agcccggtag agcccgagctc agcgcgccgc agccttcgac gcg atg ttc cgc cgg 535
 Met Phe Arg Arg
 1

agc ttg aat cgt ttt tgt gct gga gaa gag aaa cga gtt ggc aca cgc 583
 Ser Leu Asn Arg Phe Cys Ala Gly Glu Glu Lys Arg Val Gly Thr Arg
 5 10 15 20

aca gtg ttt gtt ggc aat cat cca gtt tgc gaa aca gaa gct tac att 631
 Thr Val Phe Val Gly Asn His Pro Val Ser Glu Thr Glu Ala Tyr Ile
 25 30 35

gca caa aga ttt tgt gat aat aga ata gtc tca tct aag tat aca ctt 679
 Ala Gln Arg Phe Cys Asp Asn Arg Ile Val Ser Ser Lys Tyr Thr Leu
 40 45 50

tgg aat ttt ctc cca aag aat ctg ttt gaa cag ttt aga aga att gca 727
 Trp Asn Phe Leu Pro Lys Asn Leu Phe Glu Gln Phe Arg Arg Ile Ala
 55 60 65

aat ttt tat ttt ctc ata atc ttc ctt gta cag gtc aca gta gac aca 775
 Asn Phe Tyr Phe Leu Ile Ile Phe Leu Val Gln Val Thr Val Asp Thr
 70 75 80

cca act agc cca gtt acc agt gga ctt cca ctt ttc ttt gtt ata act 823
 Pro Thr Ser Pro Val Thr Ser Gly Leu Pro Leu Phe Phe Val Ile Thr
 85 90 95 100

gtt aca gcc atc aag cag gga tat gag gat tgg ctg aga cac aga gct 871
 Val Thr Ala Ile Lys Gln Gly Tyr Glu Asp Trp Leu Arg His Arg Ala
 105 110 115

gac aat gaa gtc aac aaa agc act gtt tac att att gaa aat gca aag 919
 Asp Asn Glu Val Asn Lys Ser Thr Val Tyr Ile Ile Glu Asn Ala Lys
 120 125 130

cga gtg aga aaa gaa agt gaa aaa atc aag gtt ggt gat gta gta gaa 967
 Arg Val Arg Lys Glu Ser Glu Lys Ile Lys Val Gly Asp Val Val Glu
 135 140 145

gta cag gca gat gaa acc ttt ccc tgt gat ctt att ctt cta tca tct 1015
 Val Gln Ala Asp Glu Thr Phe Pro Cys Asp Leu Ile Leu Leu Ser Ser
 150 155 160

tgc acc act gat gga acc tgt tat gtc act aca gcc agt ctt gat ggg 1063
 Cys Thr Thr Asp Gly Thr Cys Tyr Val Thr Ala Ser Leu Asp Gly
 165 170 175 180

gaa tcc aat tgc aag aca cat tat gca gta cgt gat acc att gca ctg	1111
Glu Ser Asn Cys Lys Thr His Tyr Ala Val Arg Asp Thr Ile Ala Leu	
185 190 195	
tgt aca gca gaa tcc atc gat acc ctc cga gca gca att gaa tgt gaa	1159
Cys Thr Ala Glu Ser Ile Asp Thr Leu Arg Ala Ala Ile Glu Cys Glu	
200 205 210	
cag cct caa cct gac ctc tac aaa ttt gtt ggg cga atc aat atc tac	1207
Gln Pro Gln Pro Asp Leu Tyr Lys Phe Val Gly Arg Ile Asn Ile Tyr	
215 220 225	
agt aat agt ctt gag gct gtt gcc agg tct ttg gga cct gaa aat ctc	1255
Ser Asn Ser Leu Glu Ala Val Ala Arg Ser Leu Gly Pro Glu Asn Leu	
230 235 240	
ttg ctg aaa gga gct acg cta aaa aat acc gag aag ata tat gga gtt	1303
Leu Leu Lys Gly Ala Thr Leu Lys Asn Thr Glu Lys Ile Tyr Gly Val	
245 250 255 260	
gct gtt tac act gga atg gaa acc aaa atg gct ttg aac tac caa ggg	1351
Ala Val Tyr Thr Gly Met Glu Thr Lys Met Ala Leu Asn Tyr Gln Gly	
265 270 275	
aaa tct cag aaa cgt tct gct gtt gaa aaa tct att aat gct ttc ctg	1399
Lys Ser Gln Lys Arg Ser Ala Val Glu Lys Ser Ile Asn Ala Phe Leu	
280 285 290	
att gta tat tta ttt atc tta ctg acc aaa gct gca gta tgc act act	1447
Ile Val Tyr Leu Phe Ile Leu Leu Thr Lys Ala Ala Val Cys Thr Thr	
295 300 305	
cta aag tat gtt tgg caa agt acc cca tac aat gat gaa cct tgg tat	1495
Leu Lys Tyr Val Trp Gln Ser Thr Pro Tyr Asn Asp Glu Pro Trp Tyr	
310 315 320	
aac caa aag act cag aaa gag cga gag acc ttg aag gtt tta aaa atg	1543
Asn Gln Lys Thr Gln Lys Glu Arg Glu Thr Leu Lys Val Leu Lys Met	
325 330 335 340	
ttc acc gac ttc cta tca ttt atg gtt cta ttc aac ttt atc att cct	1591
Phe Thr Asp Phe Leu Ser Phe Met Val Leu Phe Asn Phe Ile Ile Pro	
345 350 355	
gtc tcc atg tac gtc aca gta gaa atg cag aaa ttc ttg ggc tcc ttc	1639
Val Ser Met Tyr Val Thr Val Glu Met Gln Lys Phe Leu Gly Ser Phe	
360 365 370	
ttc atc tca tgg gat aag gac ttt tat gat gaa gaa att aat gaa gga	1687
Phe Ile Ser Trp Asp Lys Asp Phe Tyr Asp Glu Glu Ile Asn Glu Gly	
375 380 385	
gcc ctg gtt aac aca tca gac ctt aat gaa gaa ctt ggt cag gtg gat	1735
Ala Leu Val Asn Thr Ser Asp Leu Asn Glu Glu Leu Gly Gln Val Asp	
390 395 400	
tat gta ttt aca gat aag act gga aca ctc act gaa aac agc atg gaa	1783
Tyr Val Phe Thr Asp Lys Thr Gly Thr Leu Thr Glu Asn Ser Met Glu	
405 410 415 420	

ttc att gaa tgc tgc ata gat ggc cac aaa tat aaa ggt gta act caa	1831
Phe Ile Glu Cys Cys Ile Asp Gly His Lys Tyr Lys Gly Val Thr Gln	
425 430 435	
gag gtt gat gga tta tct caa act gat gga act tta aca tat ttt gac	1879
Glu Val Asp Gly Leu Ser Gln Thr Asp Gly Thr Leu Thr Tyr Phe Asp	
440 445 450	
aaa gta gat aag aat cga gaa gag ctg ttt cta cgt gcc ttg tgt tta	1927
Lys Val Asp Lys Asn Arg Glu Glu Leu Phe Leu Arg Ala Leu Cys Leu	
455 460 465	
tgt cat act gta gaa atc aaa aca aac gat gct gtt gat gga gct aca	1975
Cys His Thr Val Glu Ile Lys Thr Asn Asp Ala Val Asp Gly Ala Thr	
470 475 480	
gaa tca gct gaa tta acc tat atc tcc tct tca cca gat gaa ata gct	2023
Glu Ser Ala Glu Leu Thr Tyr Ile Ser Ser Ser Pro Asp Glu Ile Ala	
485 490 495 500	
ttg gtg aaa gga gct aaa agg tac ggg ttc aca ttt tta gga aat cga	2071
Leu Val Lys Gly Ala Lys Arg Tyr Gly Phe Thr Phe Leu Gly Asn Arg	
505 510 515	
aat gga tat atg aga gta gag aac caa aga aaa gaa ata gaa gaa tat	2119
Asn Gly Tyr Met Arg Val Glu Asn Gln Arg Lys Glu Ile Glu Glu Tyr	
520 525 530	
gaa ctt ctt cac acc tta aac ttt gat gct gtc cgg cga cgt atg agt	2167
Glu Leu Leu His Thr Leu Asn Phe Asp Ala Val Arg Arg Arg Met Ser	
535 540 545	
gta att gtg aag act caa gaa gga gac ata ctt ctc ttt tgt aaa gga	2215
Val Ile Val Lys Thr Gln Glu Gly Asp Ile Leu Leu Phe Cys Lys Gly	
550 555 560	
gca gac tgc gca gtt ttt ccc aga gtg caa aat cat gaa att gag tta	2263
Ala Asp Ser Ala Val Phe Pro Arg Val Gln Asn His Glu Ile Glu Leu	
565 570 575 580	
act aaa gtc cat gtg gaa cgt aat gca atg gat ggg tat cgg aca ctc	2311
Thr Lys Val His Val Glu Arg Asn Ala Met Asp Gly Tyr Arg Thr Leu	
585 590 595	
tgt gta gcc ttc aaa gaa att gct cca gat gat tat gaa aga att aac	2359
Cys Val Ala Phe Lys Glu Ile Ala Pro Asp Asp Tyr Glu Arg Ile Asn	
600 605 610	
aga cag ctc ata gag gca aaa atg gcc tta caa gac aga gaa gaa aaa	2407
Arg Gln Leu Ile Glu Ala Lys Met Ala Leu Gln Asp Arg Glu Glu Lys	
615 620 625	
atg gaa aaa gtt ttc gat gat att gag aca aac atg aat tta att gga	2455
Met Glu Lys Val Phe Asp Asp Ile Glu Thr Asn Met Asn Leu Ile Gly	
630 635 640	
gcc act gca gtt gaa gac aag cta caa gat caa gct gca gag acc att	2503
Ala Thr Ala Val Glu Asp Lys Leu Gln Asp Gln Ala Ala Glu Thr Ile	
645 650 655 660	

gaa gct ctg cat gca gca ggc ctg aaa gtc tgg gtg ctc act ggg gac	2551
Glu Ala Leu His Ala Ala Gly Leu Lys Val Trp Val Leu Thr Gly Asp	
665 670 675	
aag atg gag aca gct aaa tcc aca tgc tat gcc tgc cgc ctt ttc cag	2599
Lys Met Glu Thr Ala Lys Ser Thr Cys Tyr Ala Cys Arg Leu Phe Gln	
680 685 690	
acc aac act gag ctc tta gaa cta acc aca aaa acc att gaa gaa agt	2647
Thr Asn Thr Glu Leu Leu Glu Thr Thr Lys Thr Ile Glu Glu Ser	
695 700 705	
gaa agg aaa gaa gat cga tta cat gaa tta ttg ata gaa tat cgc aag	2695
Glu Arg Lys Glu Asp Arg Leu His Glu Leu Leu Ile Glu Tyr Arg Lys	
710 715 720	
aaa ttg ctg cat gag ttt cct aaa agt act aga agc ttt aaa aaa gca	2743
Lys Leu Leu His Glu Phe Pro Lys Ser Thr Arg Ser Phe Lys Lys Ala	
725 730 735 740	
tgg aca gaa cat cag gaa tat gga tta atc ata gat ggc tcc aca ttg	2791
Trp Thr Glu His Gln Glu Tyr Gly Leu Ile Ile Asp Gly Ser Thr Leu	
745 750 755	
tca ctc ata cta aat tct agt caa gac tct agt tca aac aat tac aaa	2839
Ser Leu Ile Leu Asn Ser Ser Gln Asp Ser Ser Ser Asn Asn Tyr Lys	
760 765 770	
agc att ttc cta caa ata tgt atg aag tgt act gca gtg ctc tgc tgt	2887
Ser Ile Phe Leu Gln Ile Cys Met Lys Cys Thr Ala Val Leu Cys Cys	
775 780 785	
cgg atg gca cca tta cag aaa gcc cag att gtc aga atg gtg aag aat	2935
Arg Met Ala Pro Leu Gln Lys Ala Gln Ile Val Arg Met Val Lys Asn	
790 795 800	
tta aaa ggc agc cca ata act ctg tgc ata ggt gat ggt gcc aat gat	2983
Leu Lys Gly Ser Pro Ile Thr Leu Ser Ile Gly Asp Gly Ala Asn Asp	
805 810 815 820	
gtt agt atg atc ttg gaa tcc cat gtg gga ata ggt att aaa ggc aaa	3031
Val Ser Met Ile Leu Glu Ser His Val Gly Ile Gly Ile Lys Gly Lys	
825 830 835	
gaa ggt cgc caa gca gct agg aat agc gat tat tct gtt cca aag ttt	3079
Glu Gly Arg Gln Ala Ala Arg Asn Ser Asp Tyr Ser Val Pro Lys Phe	
840 845 850	
aaa cac tta aag aaa ctg ctg ttg gct cat gga cat cta tat tat gtg	3127
Lys His Leu Lys Lys Leu Leu Ala His Gly His Leu Tyr Tyr Val	
855 860 865	
aga ata gca cac ctt gta cag tac ttc ttc tat aag aac ctt tgt ttc	3175
Arg Ile Ala His Leu Val Gln Tyr Phe Phe Tyr Lys Asn Leu Cys Phe	
870 875 880	
att ttg cca cag ttt ttg tac cag ttc ttc tgt gga ttc tca caa cag	3223
Ile Leu Pro Gln Phe Leu Tyr Gln Phe Phe Cys Gly Phe Ser Gln Gln	
885 890 895 900	

cca ctg tat gat gct gct tac ctt aca atg tac aat atc tgc ttc aca 3271
Pro Leu Tyr Asp Ala Ala Tyr Leu Thr Met Tyr Asn Ile Cys Phe Thr
905 910 915

tcc ttg ccc atc ctg gcc tat agt cta ctg gaa cag cac atc aac att 3319
Ser Leu Pro Ile Leu Ala Tyr Ser Leu Leu Glu Gln His Ile Asn Ile
920 925 930

gac act ctg acc tca gat ccc cga ttg tat atg aaa att tct ggc aat 3367
Asp Thr Leu Thr Ser Asp Pro Arg Leu Tyr Met Lys Ile Ser Gly Asn
935 940 945

gcc atg cta cag ttg ggc ccc ttc tta tat tgg aca ttt ctg gct gcc 3415
Ala Met Leu Gln Leu Gly Pro Phe Leu Tyr Trp Thr Phe Leu Ala Ala
950 955 960

ttt gaa ggg aca gtg ttc ttc ttt ggg act tac ttt ctt ttt cag act 3463
Phe Glu Gly Thr Val Phe Phe Phe Gly Thr Tyr Phe Leu Phe Gln Thr
965 970 975 980

gca tcc cta gaa gaa aat gga aag gta tac gga aac tgg act ttt gga 3511
Ala Ser Leu Glu Glu Asn Gly Lys Val Tyr Gly Asn Trp Thr Phe Gly
985 990 995

acc att gtt ttt aca gtc tta gta ttc act gta acc ctg aag ctt gcc 3559
Thr Ile Val Phe Thr Val Leu Val Phe Thr Val Thr Leu Lys Leu Ala
1000 1005 1010

ttg gat acc cga ttc tgg acg tgg ata aat cac ttt gtg att tgg ggt 3607
Leu Asp Thr Arg Phe Trp Thr Trp Ile Asn His Phe Val Ile Trp Gly
1015 1020 1025

tct tta gcc ttc tat gta ttt ttc tca ttc ttc tgg gga gga att att 3655
Ser Leu Ala Phe Tyr Val Phe Phe Ser Phe Phe Trp Gly Gly Ile Ile
1030 1035 1040

tgg cct ttt ctc aag caa cag aga atg tat ttt gta ttt gcc caa atg 3703
Trp Pro Phe Leu Lys Gln Gln Arg Met Tyr Phe Val Phe Ala Gln Met
1045 1050 1055 1060

ctg tct tct gta tcc aca tgg ttg gct ata att ctt cta ata ttt atc 3751
Leu Ser Ser Val Ser Thr Trp Leu Ala Ile Ile Leu Leu Ile Phe Ile
1065 1070 1075

agc ctg ttc cct gag att ctt ctg ata gta tta aag aat gta aga aga 3799
Ser Leu Phe Pro Glu Ile Leu Leu Ile Val Leu Lys Asn Val Arg Arg
1080 1085 1090

aga agt gcc agg aga aat ctg agc tgt aga agg gca tct gac tca tta 3847
Arg Ser Ala Arg Arg Asn Leu Ser Cys Arg Arg Ala Ser Asp Ser Leu
1095 1100 1105

tcc gcc aga cct tca gtc aga cct ctt ctt tta cga aca ttc tca gac 3895
Ser Ala Arg Pro Ser Val Arg Pro Leu Leu Leu Arg Thr Phe Ser Asp
1110 1115 1120

gaa tct aat gta ttg taa cagaatccga atcttgaact gcctatgtta 3943
Glu Ser Asn Val Leu *
1125

```

ttgtcctaca agcataactga cagtgggttac agctaaaaaaa gaaagcatga agaaacaact 4003
acaaaaagtt atcatctcag gatacttgat atgcaacaca ctaaaccact ctcattgtcta 4063
gagttcacaa taaatgttca ttaaaatacc aaatgattct ctttaagcatt taccattatt 4123
gtaagtagcc tttatggcca aagctgtaag ttaagaatta tatgaaagtt gaaagcaaga 4183
atacttagaa ttctggcttt agtttagagta atataactca aatgggtgct cttttaaccc 4243
atgaactttg tgaatggatt taaatacaat agtatgaagt agaagttatg caatgagaat 4303
gaatagattt tgctaatact actttttttg cctggcagaa gaaatagact atttggatca 4363
catttctcat tctcctaaa tgatcatctt aatttttttt cccaagtaca taaggaatac 4423
ttgaaaatac agaataacta aatagtatca atgcatcaga cagaatagtt aatcccttct 4483
gtttacccat gtgctactaa tgtcttggtg gaatatctt gccaaaaaaa taccttgaac 4543
gcttatgttg aaagtgttaa cttacgggta tttttgtggg aatagaaaaa aattgtttat 4603
tttttattct tctgaattaa accccactta tgggtgtaag cctactagac ttgaaaataa 4663
agtataaac atttccaatc acttagtagc cctcaaatg agttagaaaa taaacagatt 4723
tttccagtgt tgatcttact gggatctgca gtaagggtgt ttaaaccata gttatataaa 4783
aataaaggtc attctgaata tcagcctttt ataattttat gtgaagagga agaaatatag 4843
cttattttta acttttgacg gtttttattt gaaagagatt gcatttatgc atatatgcag 4903
tgctttttct taaacttggc caatttggaa agggggaagg agccacccca aaacggtgtg 4963
tcagcttgta gagccatgac tctgtgaaga tgaatgttgt ctcttaactt ggacagggaa 5023
atggtctaac tctaaaccat gtaactgacc ttagtaaatg ccttgactaa ctgaactaga 5083
aggaagggtt agccttctaa ttagttcact tgaacataa atgtgaaatg tcttcattca 5143
atgttaaaca catacttttt tggaataaa tgaccatatt tatttgactg ctagtttttt 5203
tgtttttttt ttgtctttct ggcatgctg tactattatt aatgtttata ttgtaccttg 5263
atttgaaaaa gtattggagt taatctgtat tatatttata tagtccatat ggcacatttg 5323
attctccac atatattttg tgtaaatgtt taggtatgat tttttctaa attctagaaa 5383
agaacataat ttcagttatc agaagccatt ccatcattat agaccctttt tcattatttc 5443
atttgcctc atatatcagt attatttttg agcattttgt tacatgtcat tcacaactta 5503
cctaagtgtg ctgtgttctg gtagcccgta tttgaggtaa gctgotgaaa acaaaagtct 5563
ctatatctt tgcctattcc aaagagctaa aaaagtctaa cccaggaaaag cttttgatat 5623
tttgtgtttg ttttctgtt cttatggtt ttgttgctgt attatgattg ctgttttaca 5683
taaaatctat gggaactgtg aatacagaca agagagccac agtagagagg cttgtttaat 5743
gcagtacat tgagaggtta acagaataat ctagttagaa aataactggt tgcattgtaa 5803
attccttcca gccagaaaga aagaaagaca aggagtaagg gggatttaga gttatgtctc 5863
agctacacat tacattgtga tactgcagct caaattcaga atggcaatga tacatgatat 5923
catggcctag atccttgaga gggacctggc tttccttttt aaaagatatt ttactgaaga 5983
gctaaaaact ggccagtgtg gggttagcag atcgaataac ttgaaataga ccgtgcagta 6043
ttcctagcac tcaatgtaat caccctattt gtgacagaga aagggaaaaa aatataataa 6103
gatcatctac ctataatttg aataattttg agctatcaaa atgtctttgt aattttcaca 6163
acpgctgtcc attgtttgag gatgttacct actaaactga aaacattcat tccatatcta 6223
cttacacata caccagcaac agtataaatg taagcctaac tttgcaaaat tcgtaataat 6283
ttagtgatgg aattttttta taacatgcag tatataaatg tgcagatttt atgcgtgttg 6343
acaaaatcat ttttcagctt gcaaaatggg actgcaatat tacatttttc acttaagcag 6403
ttttttcat ctacgttggt gctttctaaa atgaatgtga atgccatctt ttatgactgc 6463
aacttgcctt ttccattaca gaaatttttg tttgatgtaa tcaataaact ttggtatgat 6523
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 6582

```

<210> 14

<211> 1129

<212> PRT

<213> Homo sapiens

<400> 14

```

Met Phe Arg Arg Ser Leu Asn Arg Phe Cys Ala Gly Glu Glu Lys Arg
 1             5             10            15
Val Gly Thr Arg Thr Val Phe Val Gly Asn His Pro Val Ser Glu Thr
             20            25            30
Glu Ala Tyr Ile Ala Gln Arg Phe Cys Asp Asn Arg Ile Val Ser Ser
             35            40            45
Lys Tyr Thr Leu Trp Asn Phe Leu Pro Lys Asn Leu Phe Glu Gln Phe
             50            55            60
Arg Arg Ile Ala Asn Phe Tyr Phe Leu Ile Ile Phe Leu Val Gln Val
             65            70            75            80

```

Thr	Val	Asp	Thr	Pro	Thr	Ser	Pro	Val	Thr	Ser	Gly	Leu	Pro	Leu	Phe
				85					90					95	
Phe	Val	Ile	Thr	Val	Thr	Ala	Ile	Lys	Gln	Gly	Tyr	Glu	Asp	Trp	Leu
			100					105					110		
Arg	His	Arg	Ala	Asp	Asn	Glu	Val	Asn	Lys	Ser	Thr	Val	Tyr	Ile	Ile
			115				120					125			
Glu	Asn	Ala	Lys	Arg	Val	Arg	Lys	Glu	Ser	Glu	Lys	Ile	Lys	Val	Gly
			130				135				140				
Asp	Val	Val	Glu	Val	Gln	Ala	Asp	Glu	Thr	Phe	Pro	Cys	Asp	Leu	Ile
					150					155					160
Leu	Leu	Ser	Ser	Cys	Thr	Thr	Asp	Gly	Thr	Cys	Tyr	Val	Thr	Thr	Ala
				165					170					175	
Ser	Leu	Asp	Gly	Glu	Ser	Asn	Cys	Lys	Thr	His	Tyr	Ala	Val	Arg	Asp
			180					185					190		
Thr	Ile	Ala	Leu	Cys	Thr	Ala	Glu	Ser	Ile	Asp	Thr	Leu	Arg	Ala	Ala
			195				200					205			
Ile	Glu	Cys	Glu	Gln	Pro	Gln	Pro	Asp	Leu	Tyr	Lys	Phe	Val	Gly	Arg
			210				215				220				
Ile	Asn	Ile	Tyr	Ser	Asn	Ser	Leu	Glu	Ala	Val	Ala	Arg	Ser	Leu	Gly
					230					235					240
Pro	Glu	Asn	Leu	Leu	Leu	Lys	Gly	Ala	Thr	Leu	Lys	Asn	Thr	Glu	Lys
				245					250					255	
Ile	Tyr	Gly	Val	Ala	Val	Tyr	Thr	Gly	Met	Glu	Thr	Lys	Met	Ala	Leu
			260					265					270		
Asn	Tyr	Gln	Gly	Lys	Ser	Gln	Lys	Arg	Ser	Ala	Val	Glu	Lys	Ser	Ile
			275				280					285			
Asn	Ala	Phe	Leu	Ile	Val	Tyr	Leu	Phe	Ile	Leu	Leu	Thr	Lys	Ala	Ala
			290				295				300				
Val	Cys	Thr	Thr	Leu	Lys	Tyr	Val	Trp	Gln	Ser	Thr	Pro	Tyr	Asn	Asp
					310					315					320
Glu	Pro	Trp	Tyr	Asn	Gln	Lys	Thr	Gln	Lys	Glu	Arg	Glu	Thr	Leu	Lys
				325					330					335	
Val	Leu	Lys	Met	Phe	Thr	Asp	Phe	Leu	Ser	Phe	Met	Val	Leu	Phe	Asn
			340					345					350		
Phe	Ile	Ile	Pro	Val	Ser	Met	Tyr	Val	Thr	Val	Glu	Met	Gln	Lys	Phe
			355				360					365			
Leu	Gly	Ser	Phe	Phe	Ile	Ser	Trp	Asp	Lys	Asp	Phe	Tyr	Asp	Glu	Glu
			370				375				380				
Ile	Asn	Glu	Gly	Ala	Leu	Val	Asn	Thr	Ser	Asp	Leu	Asn	Glu	Glu	Leu
					390					395					400
Gly	Gln	Val	Asp	Tyr	Val	Phe	Thr	Asp	Lys	Thr	Gly	Thr	Leu	Thr	Glu
				405					410					415	
Asn	Ser	Met	Glu	Phe	Ile	Glu	Cys	Cys	Ile	Asp	Gly	His	Lys	Tyr	Lys
			420					425					430		
Gly	Val	Thr	Gln	Glu	Val	Asp	Gly	Leu	Ser	Gln	Thr	Asp	Gly	Thr	Leu
			435				440					445			
Thr	Tyr	Phe	Asp	Lys	Val	Asp	Lys	Asn	Arg	Glu	Glu	Leu	Phe	Leu	Arg
			450				455				460				
Ala	Leu	Cys	Leu	Cys	His	Thr	Val	Glu	Ile	Lys	Thr	Asn	Asp	Ala	Val
					470					475					480
Asp	Gly	Ala	Thr	Glu	Ser	Ala	Glu	Leu	Thr	Tyr	Ile	Ser	Ser	Ser	Pro
				485					490					495	
Asp	Glu	Ile	Ala	Leu	Val	Lys	Gly	Ala	Lys	Arg	Tyr	Gly	Phe	Thr	Phe
			500					505					510		
Leu	Gly	Asn	Arg	Asn	Gly	Tyr	Met	Arg	Val	Glu	Asn	Gln	Arg	Lys	Glu
			515				520					525			
Ile	Glu	Glu	Tyr	Glu	Leu	Leu	His	Thr	Leu	Asn	Phe	Asp	Ala	Val	Arg
			530				535				540				
Arg	Arg	Met	Ser	Val	Ile	Val	Lys	Thr	Gln	Glu	Gly	Asp	Ile	Leu	Leu
					550					555					560

Phe	Cys	Lys	Gly	Ala	Asp	Ser	Ala	Val	Phe	Pro	Arg	Val	Gln	Asn	His	
				565					570					575		
Glu	Ile	Glu	Leu	Thr	Lys	Val	His	Val	Glu	Arg	Asn	Ala	Met	Asp	Gly	
			580					585					590			
Tyr	Arg	Thr	Leu	Cys	Val	Ala	Phe	Lys	Glu	Ile	Ala	Pro	Asp	Asp	Tyr	
		595					600					605				
Glu	Arg	Ile	Asn	Arg	Gln	Leu	Ile	Glu	Ala	Lys	Met	Ala	Leu	Gln	Asp	
	610				615						620					
Arg	Glu	Glu	Lys	Met	Glu	Lys	Val	Phe	Asp	Asp	Ile	Glu	Thr	Asn	Met	
	625				630					635					640	
Asn	Leu	Ile	Gly	Ala	Thr	Ala	Val	Glu	Asp	Lys	Leu	Gln	Asp	Gln	Ala	
			645						650					655		
Ala	Glu	Thr	Ile	Glu	Ala	Leu	His	Ala	Ala	Gly	Leu	Lys	Val	Trp	Val	
		660						665					670			
Leu	Thr	Gly	Asp	Lys	Met	Glu	Thr	Ala	Lys	Ser	Thr	Cys	Tyr	Ala	Cys	
		675					680					685				
Arg	Leu	Phe	Gln	Thr	Asn	Thr	Glu	Leu	Leu	Glu	Leu	Thr	Thr	Lys	Thr	
	690				695						700					
Ile	Glu	Glu	Ser	Glu	Arg	Lys	Glu	Asp	Arg	Leu	His	Glu	Leu	Leu	Ile	
	705				710					715					720	
Glu	Tyr	Arg	Lys	Lys	Leu	Leu	His	Glu	Phe	Pro	Lys	Ser	Thr	Arg	Ser	
			725						730					735		
Phe	Lys	Lys	Ala	Trp	Thr	Glu	His	Gln	Glu	Tyr	Gly	Leu	Ile	Ile	Asp	
			740					745					750			
Gly	Ser	Thr	Leu	Ser	Leu	Ile	Leu	Asn	Ser	Ser	Gln	Asp	Ser	Ser	Ser	
		755					760					765				
Asn	Asn	Tyr	Lys	Ser	Ile	Phe	Leu	Gln	Ile	Cys	Met	Lys	Cys	Thr	Ala	
	770				775						780					
Val	Leu	Cys	Cys	Arg	Met	Ala	Pro	Leu	Gln	Lys	Ala	Gln	Ile	Val	Arg	
	785				790					795					800	
Met	Val	Lys	Asn	Leu	Lys	Gly	Ser	Pro	Ile	Thr	Leu	Ser	Ile	Gly	Asp	
			805						810					815		
Gly	Ala	Asn	Asp	Val	Ser	Met	Ile	Leu	Glu	Ser	His	Val	Gly	Ile	Gly	
		820						825					830			
Ile	Lys	Gly	Lys	Glu	Gly	Arg	Gln	Ala	Ala	Arg	Asn	Ser	Asp	Tyr	Ser	
		835					840					845				
Val	Pro	Lys	Phe	Lys	His	Leu	Lys	Lys	Leu	Leu	Leu	Ala	His	Gly	His	
	850					855					860					
Leu	Tyr	Tyr	Val	Arg	Ile	Ala	His	Leu	Val	Gln	Tyr	Phe	Phe	Tyr	Lys	
	865				870					875					880	
Asn	Leu	Cys	Phe	Ile	Leu	Pro	Gln	Phe	Leu	Tyr	Gln	Phe	Phe	Cys	Gly	
			885					890						895		
Phe	Ser	Gln	Gln	Pro	Leu	Tyr	Asp	Ala	Ala	Tyr	Leu	Thr	Met	Tyr	Asn	
			900					905					910			
Ile	Cys	Phe	Thr	Ser	Leu	Pro	Ile	Leu	Ala	Tyr	Ser	Leu	Leu	Glu	Gln	
		915					920					925				
His	Ile	Asn	Ile	Asp	Thr	Leu	Thr	Ser	Asp	Pro	Arg	Leu	Tyr	Met	Lys	
		930					935				940					
Ile	Ser	Gly	Asn	Ala	Met	Leu	Gln	Leu	Gly	Pro	Phe	Leu	Tyr	Trp	Thr	
	945				950					955					960	
Phe	Leu	Ala	Ala	Phe	Glu	Gly	Thr	Val	Phe	Phe	Phe	Gly	Thr	Tyr	Phe	
			965					970						975		
Leu	Phe	Gln	Thr	Ala	Ser	Leu	Glu	Glu	Asn	Gly	Lys	Val	Tyr	Gly	Asn	
			980					985					990			
Trp	Thr	Phe	Gly	Thr	Ile	Val	Phe	Thr	Val	Leu	Val	Phe	Thr	Val	Thr	
		995					1000					1005				
Leu	Lys	Leu	Ala	Leu	Asp	Thr	Arg	Phe	Trp	Thr	Trp	Ile	Asn	His	Phe	
		1010				1015					1020					
Val	Ile	Trp	Gly	Ser	Leu	Ala	Phe	Tyr	Val	Phe	Phe	Ser	Phe	Phe	Trp	
	1025				1030					1035					1040	

```

Gly Gly Ile Ile Trp Pro Phe Leu Lys Gln Gln Arg Met Tyr Phe Val
      1045      1050      1055
Phe Ala Gln Met Leu Ser Ser Val Ser Thr Trp Leu Ala Ile Ile Leu
      1060      1065      1070
Leu Ile Phe Ile Ser Leu Phe Pro Glu Ile Leu Leu Ile Val Leu Lys
      1075      1080      1085
Asn Val Arg Arg Arg Ser Ala Arg Arg Asn Leu Ser Cys Arg Arg Ala
      1090      1095      1100
Ser Asp Ser Leu Ser Ala Arg Pro Ser Val Arg Pro Leu Leu Leu Arg
      1105      1110      1115      1120
Thr Phe Ser Asp Glu Ser Asn Val Leu
      1125

```

<210> 15
 <211> 3390
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (1)...(3390)

```

<400> 15
atg ttc cgc cgg agc ttg aat cgt ttt tgt gct gga gaa gag aaa cga      48
Met Phe Arg Arg Ser Leu Asn Arg Phe Cys Ala Gly Glu Glu Lys Arg
  1           5           10           15

ggt ggc aca cgc aca gtg ttt gtt ggc aat cat cca gtt tcg gaa aca      96
Val Gly Thr Arg Thr Val Phe Val Gly Asn His Pro Val Ser Glu Thr
           20           25           30

gaa gct tac att gca caa aga ttt tgt gat aat aga ata gtc tca tct     144
Glu Ala Tyr Ile Ala Gln Arg Phe Cys Asp Asn Arg Ile Val Ser Ser
           35           40           45

aag tat aca ctt tgg aat ttt ctc cca aag aat ctg ttt gaa cag ttt     192
Lys Tyr Thr Leu Trp Asn Phe Leu Pro Lys Asn Leu Phe Glu Gln Phe
           50           55           60

aga aga att gca aat ttt tat ttt ctc ata atc ttc ctt gta cag gtc     240
Arg Arg Ile Ala Asn Phe Tyr Phe Leu Ile Ile Phe Leu Val Gln Val
           65           70           75           80

aca gta gac aca cca act agc cca gtt acc agt gga ctt cca ctt ttc     288
Thr Val Asp Thr Pro Thr Ser Pro Val Thr Ser Gly Leu Pro Leu Phe
           85           90           95

ttt gtt ata act gtt aca gcc atc aag cag gga tat gag gat tgg ctg     336
Phe Val Ile Thr Val Thr Ala Ile Lys Gln Gly Tyr Glu Asp Trp Leu
           100          105          110

aga cac aga gct gac aat gaa gtc aac aaa agc act gtt tac att att     384
Arg His Arg Ala Asp Asn Glu Val Asn Lys Ser Thr Val Tyr Ile Ile
           115          120          125

gaa aat gca aag cga gtg aga aaa gaa agt gaa aaa atc aag gtt ggt     432
Glu Asn Ala Lys Arg Val Arg Lys Glu Ser Glu Lys Ile Lys Val Gly
           130          135          140

```

gat gta gta gaa gta cag gca gat gaa acc ttt ccc tgt gat ctt att	480
Asp Val Val Glu Val Gln Ala Asp Glu Thr Phe Pro Cys Asp Leu Ile	
145 150 155 160	
ctt cta tca tct tgc acc act gat gga acc tgt tat gtc act aca gcc	528
Leu Leu Ser Ser Cys Thr Thr Asp Gly Thr Cys Tyr Val Thr Thr Ala	
165 170 175	
agt ctt gat ggg gaa tcc aat tgc aag aca cat tat gca gta cgt gat	576
Ser Leu Asp Gly Glu Ser Asn Cys Lys Thr His Tyr Ala Val Arg Asp	
180 185 190	
acc att gca ctg tgt aca gca gaa tcc atc gat acc ctc cga gca gca	624
Thr Ile Ala Leu Cys Thr Ala Glu Ser Ile Asp Thr Leu Arg Ala Ala	
195 200 205	
att gaa tgt gaa cag cct caa cct gac ctc tac aaa ttt gtt ggg cga	672
Ile Glu Cys Glu Gln Pro Gln Pro Asp Leu Tyr Lys Phe Val Gly Arg	
210 215 220	
atc aat atc tac agt aat agt ctt gag gct gtt gcc agg tct ttg gga	720
Ile Asn Ile Tyr Ser Asn Ser Leu Glu Ala Val Ala Arg Ser Leu Gly	
225 230 235 240	
cct gaa aat ctc ttg ctg aaa gga gct acg cta aaa aat acc gag aag	768
Pro Glu Asn Leu Leu Leu Lys Gly Ala Thr Leu Lys Asn Thr Glu Lys	
245 250 255	
ata tat gga gtt gct gtt tac act gga atg gaa acc aaa atg gct ttg	816
Ile Tyr Gly Val Ala Val Tyr Thr Gly Met Glu Thr Lys Met Ala Leu	
260 265 270	
aac tac caa ggg aaa tct cag aaa cgt tct gct gtt gaa aaa tct att	864
Asn Tyr Gln Gly Lys Ser Gln Lys Arg Ser Ala Val Glu Lys Ser Ile	
275 280 285	
aat gct ttc ctg att gta tat tta ttt atc tta ctg acc aaa gct gca	912
Asn Ala Phe Leu Ile Val Tyr Leu Phe Ile Leu Leu Thr Lys Ala Ala	
290 295 300	
gta tgc act act cta aag tat gtt tgg caa agt acc cca tac aat gat	960
Val Cys Thr Thr Leu Lys Tyr Val Trp Gln Ser Thr Pro Tyr Asn Asp	
305 310 315 320	
gaa cct tgg tat aac caa aag act cag aaa gag cga gag acc ttg aag	1008
Glu Pro Trp Tyr Asn Gln Lys Thr Gln Lys Glu Arg Glu Thr Leu Lys	
325 330 335	
gtt tta aaa atg ttc acc gac ttc cta tca ttt atg gtt cta ttc aac	1056
Val Leu Lys Met Phe Thr Asp Phe Leu Ser Phe Met Val Leu Phe Asn	
340 345 350	
ttt atc att cct gtc tcc atg tac gtc aca gta gaa atg cag aaa ttc	1104
Phe Ile Ile Pro Val Ser Met Tyr Val Thr Val Glu Met Gln Lys Phe	
355 360 365	
ttg ggc tcc ttc ttc atc tca tgg gat aag gac ttt tat gat gaa gaa	1152
Leu Gly Ser Phe Phe Ile Ser Trp Asp Lys Asp Phe Tyr Asp Glu Glu	
370 375 380	

att aat gaa gga gcc ctg gtt aac aca tca gac ctt aat gaa gaa ctt	1200
Ile Asn Glu Gly Ala Leu Val Asn Thr Ser Asp Leu Asn Glu Glu Leu	
385 390 395 400	
ggt cag gtg gat tat gta ttt aca gat aag act gga aca ctc act gaa	1248
Gly Gln Val Asp Tyr Val Phe Thr Asp Lys Thr Gly Thr Leu Thr Glu	
405 410 415	
aac agc atg gaa ttc att gaa tgc tgc ata gat ggc cac aaa tat aaa	1296
Asn Ser Met Glu Phe Ile Glu Cys Cys Ile Asp Gly His Lys Tyr Lys	
420 425 430	
ggt gta act caa gag gtt gat gga tta tct caa act gat gga act tta	1344
Gly Val Thr Gln Glu Val Asp Gly Leu Ser Gln Thr Asp Gly Thr Leu	
435 440 445	
aca tat ttt gac aaa gta gat aag aat cga gaa gag ctg ttt cta cgt	1392
Thr Tyr Phe Asp Lys Val Asp Lys Asn Arg Glu Glu Leu Phe Leu Arg	
450 455 460	
gcc ttg tgt tta tgt cat act gta gaa atc aaa aca aac gat gct gtt	1440
Ala Leu Cys Leu Cys His Thr Val Glu Ile Lys Thr Asn Asp Ala Val	
465 470 475 480	
gat gga gct aca gaa tca gct gaa tta acc tat atc tcc tct tca cca	1488
Asp Gly Ala Thr Glu Ser Ala Glu Leu Thr Tyr Ile Ser Ser Ser Pro	
485 490 495	
gat gaa ata gct ttg gtg aaa gga gct aaa agg tac ggg ttc aca ttt	1536
Asp Glu Ile Ala Leu Val Lys Gly Ala Lys Arg Tyr Gly Phe Thr Phe	
500 505 510	
tta gga aat cga aat gga tat atg aga gta gag aac caa aga aaa gaa	1584
Leu Gly Asn Arg Asn Gly Tyr Met Arg Val Glu Asn Gln Arg Lys Glu	
515 520 525	
ata gaa gaa tat gaa ctt ctt cac acc tta aac ttt gat gct gtc cgg	1632
Ile Glu Glu Tyr Glu Leu Leu His Thr Leu Asn Phe Asp Ala Val Arg	
530 535 540	
cga cgt atg agt gta att gtg aag act caa gaa gga gac ata ctt ctc	1680
Arg Arg Met Ser Val Ile Val Lys Thr Gln Glu Gly Asp Ile Leu Leu	
545 550 555 560	
ttt tgt aaa gga gca gac tcg gca gtt ttt ccc aga gtg caa aat cat	1728
Phe Cys Lys Gly Ala Asp Ser Ala Val Phe Pro Arg Val Gln Asn His	
565 570 575	
gaa att gag tta act aaa gtc cat gtg gaa cgt aat gca atg gat ggg	1776
Glu Ile Glu Leu Thr Lys Val His Val Glu Arg Asn Ala Met Asp Gly	
580 585 590	
tat cgg aca ctc tgt gta gcc ttc aaa gaa att gct cca gat gat tat	1824
Tyr Arg Thr Leu Cys Val Ala Phe Lys Glu Ile Ala Pro Asp Asp Tyr	
595 600 605	
gaa aga att aac aga cag ctc ata gag gca aaa atg gcc tta caa gac	1872
Glu Arg Ile Asn Arg Gln Leu Ile Glu Ala Lys Met Ala Leu Gln Asp	
610 615 620	

aga gaa gaa aaa atg gaa aaa gtt ttc gat gat att gag aca aac atg	1920
Arg Glu Glu Lys Met Glu Lys Val Phe Asp Asp Ile Glu Thr Asn Met	
625 630 635 640	
aat tta att gga gcc act gca gtt gaa gac aag cta caa gat caa gct	1968
Asn Leu Ile Gly Ala Thr Ala Val Glu Asp Lys Leu Gln Asp Gln Ala	
645 650 655	
gca gag acc att gaa gct ctg cat gca gca ggc ctg aaa gtc tgg gtg	2016
Ala Glu Thr Ile Glu Ala Leu His Ala Ala Gly Leu Lys Val Trp Val	
660 665 670	
ctc act ggg gac aag atg gag aca gct aaa tcc aca tgc tat gcc tgc	2064
Leu Thr Gly Asp Lys Met Glu Thr Ala Lys Ser Thr Cys Tyr Ala Cys	
675 680 685	
cgc ctt ttc cag acc aac act gag ctc tta gaa cta acc aca aaa acc	2112
Arg Leu Phe Gln Thr Asn Thr Glu Leu Leu Glu Leu Thr Thr Lys Thr	
690 695 700	
att gaa gaa agt gaa agg aaa gaa gat cga tta cat gaa tta ttg ata	2160
Ile Glu Glu Ser Glu Arg Lys Glu Asp Arg Leu His Glu Leu Leu Ile	
705 710 715 720	
gaa tat cgc aag aaa ttg ctg cat gag ttt cct aaa agt act aga agc	2208
Glu Tyr Arg Lys Lys Leu Leu His Glu Phe Pro Lys Ser Thr Arg Ser	
725 730 735	
ttt aaa aaa gca tgg aca gaa cat cag gaa tat gga tta atc ata gat	2256
Phe Lys Lys Ala Trp Thr Glu His Gln Glu Tyr Gly Leu Ile Ile Asp	
740 745 750	
ggc tcc aca ttg tca ctc ata cta aat tct agt caa gac tct agt tca	2304
Gly Ser Thr Leu Ser Leu Ile Leu Asn Ser Ser Gln Asp Ser Ser Ser	
755 760 765	
aac aat tac aaa agc att ttc cta caa ata tgt atg aag tgt act gca	2352
Asn Asn Tyr Lys Ser Ile Phe Leu Gln Ile Cys Met Lys Cys Thr Ala	
770 775 780	
gtg ctc tgc tgt cgg atg gca cca tta cag aaa gcc cag att gtc aga	2400
Val Leu Cys Cys Arg Met Ala Pro Leu Gln Lys Ala Gln Ile Val Arg	
785 790 795 800	
atg gtg aag aat tta aaa ggc agc cca ata act ctg tgc ata ggt gat	2448
Met Val Lys Asn Leu Lys Gly Ser Pro Ile Thr Leu Ser Ile Gly Asp	
805 810 815	
ggt gcc aat gat gtt agt atg atc ttg gaa tcc cat gtg gga ata ggt	2496
Gly Ala Asn Asp Val Ser Met Ile Leu Glu Ser His Val Gly Ile Gly	
820 825 830	
att aaa ggc aaa gaa ggt cgc caa gca gct agg aat agc gat tat tct	2544
Ile Lys Gly Lys Glu Gly Arg Gln Ala Ala Arg Asn Ser Asp Tyr Ser	
835 840 845	
gtt cca aag ttt aaa cac tta aag aaa ctg ctg ttg gct cat gga cat	2592
Val Pro Lys Phe Lys His Leu Lys Lys Leu Leu Leu Ala His Gly His	
850 855 860	

cta tat tat gtg aga ata gca cac ctt gta cag tac ttc ttc tat aag	2640
Leu Tyr Tyr Val Arg Ile Ala His Leu Val Gln Tyr Phe Phe Tyr Lys	
865 870 875 880	
aac ctt tgt ttc att ttg cca cag ttt ttg tac cag ttc ttc tgt gga	2688
Asn Leu Cys Phe Ile Leu Pro Gln Phe Leu Tyr Gln Phe Phe Cys Gly	
885 890 895	
ttc tca caa cag cca ctg tat gat gct gct tac ctt aca atg tac aat	2736
Phe Ser Gln Gln Pro Leu Tyr Asp Ala Ala Tyr Leu Thr Met Tyr Asn	
900 905 910	
atc tgc ttc aca tcc ttg ccc atc ctg gcc tat agt cta ctg gaa cag	2784
Ile Cys Phe Thr Ser Leu Pro Ile Leu Ala Tyr Ser Leu Leu Glu Gln	
915 920 925	
cac atc aac att gac act ctg acc tca gat ccc cga ttg tat atg aaa	2832
His Ile Asn Ile Asp Thr Leu Thr Ser Asp Pro Arg Leu Tyr Met Lys	
930 935 940	
att tct ggc aat gcc atg cta cag ttg ggc ccc ttc tta tat tgg aca	2880
Ile Ser Gly Asn Ala Met Leu Gln Leu Gly Pro Phe Leu Tyr Trp Thr	
945 950 955 960	
ttt ctg gct gcc ttt gaa ggg aca gtg ttc ttc ttt ggg act tac ttt	2928
Phe Leu Ala Ala Phe Glu Gly Thr Val Phe Phe Phe Gly Thr Tyr Phe	
965 970 975	
ctt ttt cag act gca tcc cta gaa gaa aat gga aag gta tac gga aac	2976
Leu Phe Gln Thr Ala Ser Leu Glu Glu Asn Gly Lys Val Tyr Gly Asn	
980 985 990	
tgg act ttt gga acc att gtt ttt aca gtc tta gta ttc act gta acc	3024
Trp Thr Phe Gly Thr Ile Val Phe Thr Val Leu Val Phe Thr Val Thr	
995 1000 1005	
ctg aag ctt gcc ttg gat acc cga ttc tgg acg tgg ata aat cac ttt	3072
Leu Lys Leu Ala Leu Asp Thr Arg Phe Trp Thr Trp Ile Asn His Phe	
1010 1015 1020	
gtg att tgg ggt tct tta gcc ttc tat gta ttt ttc tca ttc ttc tgg	3120
Val Ile Trp Gly Ser Leu Ala Phe Tyr Val Phe Phe Ser Phe Phe Trp	
1025 1030 1035 1040	
gga gga att att tgg cct ttt ctc aag caa cag aga atg tat ttt gta	3168
Gly Gly Ile Ile Trp Pro Phe Leu Lys Gln Gln Arg Met Tyr Phe Val	
1045 1050 1055	
ttt gcc caa atg ctg tct tct gta tcc aca tgg ttg gct ata att ctt	3216
Phe Ala Gln Met Leu Ser Ser Val Ser Thr Trp Leu Ala Ile Ile Leu	
1060 1065 1070	
cta ata ttt atc agc ctg ttc cct gag att ctt ctg ata gta tta aag	3264
Leu Ile Phe Ile Ser Leu Phe Pro Glu Ile Leu Leu Ile Val Leu Lys	
1075 1080 1085	
aat gta aga aga aga agt gcc agg aga aat ctg agc tgt aga agg gca	3312
Asn Val Arg Arg Arg Ser Ala Arg Arg Asn Leu Ser Cys Arg Arg Ala	
1090 1095 1100	

tct gac tca tta tcc gcc aga cct tca gtc aga cct ctt ctt tta cga 3360
 Ser Asp Ser Leu Ser Ala Arg Pro Ser Val Arg Pro Leu Leu Leu Arg
 1105 1110 1115 1120

aca ttc tca gac gaa tct aat gta ttg taa 3390
 Thr Phe Ser Asp Glu Ser Asn Val Leu *
 1125

<210> 16

<211> 6073

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (273)... (4553)

<400> 16

ccacgcgtcc gggaggagcg gagggagaag taggttgcca gctcagcaca ggctccggcg 60
 ctggctcccg cagctgagtt tgggagatgt ctaagtgatt tttttttttt ccggaaggca 120
 aatggctggc gtggaagcac aaccgcgttt cactcttcga atttgctt agctcttttc 180
 ttgtaccttg cgactcgtga ccaacatgct gtgatgtgtg ccgaggagg aattggtcag 240
 ctacacaacc tggatcttac cacagtgttg at atg act gag gct ctc caa tgg 293
 Met Thr Glu Ala Leu Gln Trp
 1 5

gcc aga tat cac tgg cga cgg ctg atc aga ggt gca acc agg gat gat 341
 Ala Arg Tyr His Trp Arg Arg Leu Ile Arg Gly Ala Thr Arg Asp Asp
 10 15 20

gat tca ggg cca tac aac tat tcc tgg ttg ctc gcc tgt ggg cgc aag 389
 Asp Ser Gly Pro Tyr Asn Tyr Ser Ser Leu Leu Ala Cys Gly Arg Lys
 25 30 35

tcc tct cag atc cct aaa ctg tca gga agg cac cgg att gtt gtt ccc 437
 Ser Ser Gln Ile Pro Lys Leu Ser Gly Arg His Arg Ile Val Val Pro
 40 45 50 55

cac atc cag ccc ttc aag gat gag tat gag aag ttc tcc gga gcc tat 485
 His Ile Gln Pro Phe Lys Asp Glu Tyr Glu Lys Phe Ser Gly Ala Tyr
 60 65 70

gtg aac aat cga ata cga aca aca aag tac aca ctt ctg aat ttt gtg 533
 Val Asn Asn Arg Ile Arg Thr Thr Lys Tyr Thr Leu Leu Asn Phe Val
 75 80 85

cca aga aat tta ttt gaa caa ttt cac aga gct gcc agt tta tat ttc 581
 Pro Arg Asn Leu Phe Glu Gln Phe His Arg Ala Ala Ser Leu Tyr Phe
 90 95 100

ctg ttc cta gtt gtc ctg aac tgg gta cct ttg gta gaa gcc ttc caa 629
 Leu Phe Leu Val Val Leu Asn Trp Val Pro Leu Val Glu Ala Phe Gln
 105 110 115

aag gaa atc acc atg ttg cct ctg gtg gtg gtc ctt aca att atc gca 677
 Lys Glu Ile Thr Met Leu Pro Leu Val Val Val Leu Thr Ile Ile Ala
 120 125 130 135

att aaa gat ggc ctg gaa gat tat cgg aaa tac aaa att gac aaa cag	725
Ile Lys Asp Gly Leu Glu Asp Tyr Arg Lys Tyr Lys Ile Asp Lys Gln	
140 145 150	
atc aat aat tta ata act aaa gtt tat agt agg aaa gag aaa aaa tac	773
Ile Asn Asn Leu Ile Thr Lys Val Tyr Ser Arg Lys Glu Lys Lys Tyr	
155 160 165	
att gac cga tgc tgg aaa gac gtt act gtt ggg gac ttt att cgc ctc	821
Ile Asp Arg Cys Trp Lys Asp Val Thr Val Gly Asp Phe Ile Arg Leu	
170 175 180	
tcc tgc aat gag gtc atc cct gca gac atg gta cta ctc ttt tcc act	869
Ser Cys Asn Glu Val Ile Pro Ala Asp Met Val Leu Leu Phe Ser Thr	
185 190 195	
gat cca gat gga atc tgt cac att gag act tct ggt ctt gat gga gag	917
Asp Pro Asp Gly Ile Cys His Ile Glu Thr Ser Gly Leu Asp Gly Glu	
200 205 210 215	
agc aat tta aaa cag agg cag gtg gtt cgg gga tat gca gaa cag gac	965
Ser Asn Leu Lys Gln Arg Gln Val Val Arg Gly Tyr Ala Glu Gln Asp	
220 225 230	
tct gaa gtt gat cct gag aag ttt tcc agt agg ata gaa tgt gaa agc	1013
Ser Glu Val Asp Pro Glu Lys Phe Ser Ser Arg Ile Glu Cys Glu Ser	
235 240 245	
cca aac aat gac ctc agc aga ttc cga ggc ttc cta gaa cat tcc aac	1061
Pro Asn Asn Asp Leu Ser Arg Phe Arg Gly Phe Leu Glu His Ser Asn	
250 255 260	
aaa gaa cgc gtg ggt ctc agt aaa gaa aat ttg ttg ctt aga gga tgc	1109
Lys Glu Arg Val Gly Leu Ser Lys Glu Asn Leu Leu Leu Arg Gly Cys	
265 270 275	
acc att aga aac aca gag gct gtt gtg ggc att gtg gtt tat gca ggc	1157
Thr Ile Arg Asn Thr Glu Ala Val Val Gly Ile Val Val Tyr Ala Gly	
280 285 290 295	
cat gaa acc aaa gca atg ctg aac aac agt ggg cca cgg tat aag cgc	1205
His Glu Thr Lys Ala Met Leu Asn Asn Ser Gly Pro Arg Tyr Lys Arg	
300 305 310	
agc aaa tta gaa aga aga gca aac aca gat gtc ctc tgg tgt gtc atg	1253
Ser Lys Leu Glu Arg Arg Ala Asn Thr Asp Val Leu Trp Cys Val Met	
315 320 325	
ctt ctg gtc ata atg tgc tta act ggc gca gta ggt cat gga atc tgg	1301
Leu Leu Val Ile Met Cys Leu Thr Gly Ala Val Gly His Gly Ile Trp	
330 335 340	
ctg agc agg tat gaa aag atg cat ttt ttc aat gtt ccc gag cct gat	1349
Leu Ser Arg Tyr Glu Lys Met His Phe Phe Asn Val Pro Glu Pro Asp	
345 350 355	
gga cat atc ata tca cca ctg ttg gca gga ttt tat atg ttt tgg acc	1397
Gly His Ile Ile Ser Pro Leu Leu Ala Gly Phe Tyr Met Phe Trp Thr	
360 365 370 375	

atg atc att ttg tta cag gtc ttg att cct att tct ctc tat gtt tcc	1445
Met Ile Ile Leu Leu Gln Val Leu Ile Pro Ile Ser Leu Tyr Val Ser	
380 385 390	
atc gaa att gtg aag ctt gga caa ata tat ttc att caa agt gat gtg	1493
Ile Glu Ile Val Lys Leu Gly Gln Ile Tyr Phe Ile Gln Ser Asp Val	
395 400 405	
gat ttc tac aat gaa aaa atg gat tct att gtt cag tgc cga gcc ctg	1541
Asp Phe Tyr Asn Glu Lys Met Asp Ser Ile Val Gln Cys Arg Ala Leu	
410 415 420	
aac atc gcc gag gat ctg gga cag att cag tac ctc ttt tcc gat aag	1589
Asn Ile Ala Glu Asp Leu Gly Gln Ile Gln Tyr Leu Phe Ser Asp Lys	
425 430 435	
acp gga acc ctc act gag aat aag atg gtt ttt cga aga tgt agt gtg	1637
Thr Gly Thr Leu Thr Glu Asn Lys Met Val Phe Arg Arg Cys Ser Val	
440 445 450 455	
gca gga ttt gat tac tgc cat gaa gaa aat gcc agg agg ttg gag tcc	1685
Ala Gly Phe Asp Tyr Cys His Glu Glu Asn Ala Arg Arg Leu Glu Ser	
460 465 470	
tat cag gaa gct gtc tct gaa gat gaa gat ttt ata gac aca gtc agt	1733
Tyr Gln Glu Ala Val Ser Glu Asp Glu Asp Phe Ile Asp Thr Val Ser	
475 480 485	
ggt tcc ctc agc aat atg gca aaa ccg aga gcc ccc agc tgc agg aca	1781
Gly Ser Leu Ser Asn Met Ala Lys Pro Arg Ala Pro Ser Cys Arg Thr	
490 495 500	
gtt cat aat ggg cct ttg gga aat aag ccc tca aat cat ctt gct ggg	1829
Val His Asn Gly Pro Leu Gly Asn Lys Pro Ser Asn His Leu Ala Gly	
505 510 515	
agc tct ttt act cta gga agt gga gaa gga gcc agt gaa gtg cct cat	1877
Ser Ser Phe Thr Leu Gly Ser Gly Glu Gly Ala Ser Glu Val Pro His	
520 525 530 535	
tcc aga cag gct gct ttc agt agc ccc att gaa aca gac gtg gta cca	1925
Ser Arg Gln Ala Ala Phe Ser Ser Pro Ile Glu Thr Asp Val Val Pro	
540 545 550	
gac acc agg ctt tta gac aaa ttt agt cag att aca cct cgg ctc ttt	1973
Asp Thr Arg Leu Leu Asp Lys Phe Ser Gln Ile Thr Pro Arg Leu Phe	
555 560 565	
atg cca cta gat gag acc atc caa aat cca cca atg gaa act ttg tac	2021
Met Pro Leu Asp Glu Thr Ile Gln Asn Pro Pro Met Glu Thr Leu Tyr	
570 575 580	
att atc gac ttt ttc att gca ttg gca att tgc aac aca gta gtg gtt	2069
Ile Ile Asp Phe Phe Ile Ala Leu Ala Ile Cys Asn Thr Val Val Val	
585 590 595	
tct gct cct aac caa ccc cga caa aag atc aga cac cct tca ctg ggg	2117
Ser Ala Pro Asn Gln Pro Arg Gln Lys Ile Arg His Pro Ser Leu Gly	
600 605 610 615	

ggg ttg ccc att aag tct ttg gaa gag att aaa agt ctt ttc cag aga	2165
Gly Leu Pro Ile Lys Ser Leu Glu Glu Ile Lys Ser Leu Phe Gln Arg	
620 625 630	
tgg tct gtc cga aga tca agt tct cca tgc ctt aac agt ggg aaa gag	2213
Trp Ser Val Arg Arg Ser Ser Ser Pro Ser Leu Asn Ser Gly Lys Glu	
635 640 645	
cca tct tct gga gtt cca aac gcc ttt gtg agc aga ctc cct ctc ttt	2261
Pro Ser Ser Gly Val Pro Asn Ala Phe Val Ser Arg Leu Pro Leu Phe	
650 655 660	
agt cga atg aaa cca gct tca cct gtg gag gaa gag gtc tcc cag gtg	2309
Ser Arg Met Lys Pro Ala Ser Pro Val Glu Glu Glu Val Ser Gln Val	
665 670 675	
tgt gag agc ccc cag tgc tcc agt agc tca gct tgc tgc aca gaa aca	2357
Cys Glu Ser Pro Gln Cys Ser Ser Ser Ser Ala Cys Cys Thr Glu Thr	
680 685 690 695	
gag aaa caa cac ggt gat gca ggc ctc ctg aat ggc aag gca gag tcc	2405
Glu Lys Gln His Gly Asp Ala Gly Leu Leu Asn Gly Lys Ala Glu Ser	
700 705 710	
ctc cct gga cag cca ttg gcc tgc aac ctg tgt tat gag gcc gag agc	2453
Leu Pro Gly Gln Pro Leu Ala Cys Asn Leu Cys Tyr Glu Ala Glu Ser	
715 720 725	
cca gac gaa gcg gcc tta gtg tat gcc gcc agg gct tac caa tgc act	2501
Pro Asp Glu Ala Ala Leu Val Tyr Ala Ala Arg Ala Tyr Gln Cys Thr	
730 735 740	
tta cgg tct cgg aca cca gag cag gtc atg gtg gac ttt gct gct ttg	2549
Leu Arg Ser Arg Thr Pro Glu Gln Val Met Val Asp Phe Ala Ala Leu	
745 750 755	
gga cca tta aca ttt caa ctc cta cac atc ctg ccc ttt gac tca gta	2597
Gly Pro Leu Thr Phe Gln Leu Leu His Ile Leu Pro Phe Asp Ser Val	
760 765 770 775	
aga aaa aga atg tct gtt gtg gtc cga cac cct ctt tcc aat caa gtt	2645
Arg Lys Arg Met Ser Val Val Val Arg His Pro Leu Ser Asn Gln Val	
780 785 790	
gtg gtg tat acg aaa ggc gct gat tct gtg atc atg gag tta ctg tgc	2693
Val Val Tyr Thr Lys Gly Ala Asp Ser Val Ile Met Glu Leu Leu Ser	
795 800 805	
gtg gct tcc cca gat gga gca agt ctg gag aaa caa cag atg ata gta	2741
Val Ala Ser Pro Asp Gly Ala Ser Leu Glu Lys Gln Gln Met Ile Val	
810 815 820	
agg gag aaa acc cag aag cac ttg gat gac tat gcc aaa caa ggc ctt	2789
Arg Glu Lys Thr Gln Lys His Leu Asp Asp Tyr Ala Lys Gln Gly Leu	
825 830 835	
cgt act tta tgt ata gca aag aag gtc atg agt gac act gaa tat gca	2837
Arg Thr Leu Cys Ile Ala Lys Lys Val Met Ser Asp Thr Glu Tyr Ala	
840 845 850 855	

gag tgg ctg agg aat cat ttt tta gct gaa acc agc att gac aac agg	2885
Glu Trp Leu Arg Asn His Phe Leu Ala Glu Thr Ser Ile Asp Asn Arg	
860 865 870	
gaa gaa tta cta ctt gaa tct gcc atg agg ttg gag aac aaa ctt aca	2933
Glu Glu Leu Leu Leu Glu Ser Ala Met Arg Leu Glu Asn Lys Leu Thr	
875 880 885	
tta ctt ggt gct act ggc att gaa gac cgt ctg cag gag gga gtc cct	2981
Leu Leu Gly Ala Thr Gly Ile Glu Asp Arg Leu Gln Glu Gly Val Pro	
890 895 900	
gaa tct ata gaa gct ctt cac aaa gcg ggc atc aag atc tgg atg ctg	3029
Glu Ser Ile Glu Ala Leu His Lys Ala Gly Ile Lys Ile Trp Met Leu	
905 910 915	
aca ggg gac aag cag gag aca gct gtc aac ata gct tat gca tgc aaa	3077
Thr Gly Asp Lys Gln Glu Thr Ala Val Asn Ile Ala Tyr Ala Cys Lys	
920 925 930 935	
cta ctg gag cca gat gac aag ctt ttt atc ctc aat acc caa agt aaa	3125
Leu Leu Glu Pro Asp Asp Lys Leu Phe Ile Leu Asn Thr Gln Ser Lys	
940 945 950	
gat gcc tgt ggg atg ctg atg agc aca att ttg aaa gaa ctt cag aag	3173
Asp Ala Cys Gly Met Leu Met Ser Thr Ile Leu Lys Glu Leu Gln Lys	
955 960 965	
aaa act caa gcc ctg cca gag caa gtg tca tta agt gaa gat tta ctt	3221
Lys Thr Gln Ala Leu Pro Glu Gln Val Ser Leu Ser Glu Asp Leu Leu	
970 975 980	
cag cct cct gtc ccc cgg gac tca ggg tta cga gct gga ctc att atc	3269
Gln Pro Pro Val Pro Arg Asp Ser Gly Leu Arg Ala Gly Leu Ile Ile	
985 990 995	
act ggg aag acc ctg gag ttt gcc ctg caa gaa agt ctg caa aag cag	3317
Thr Gly Lys Thr Leu Glu Phe Ala Leu Gln Glu Ser Leu Gln Lys Gln	
1000 1005 1010 1015	
ttc ctg gaa ctg aca tct tgg tgt caa gct gtg gtc tgc tgc cga gcc	3365
Phe Leu Glu Leu Thr Ser Trp Cys Gln Ala Val Val Cys Cys Arg Ala	
1020 1025 1030	
aca ccg ctg cag aaa agt gaa gtg gtg aaa ttg gtc cgc agc cat ctc	3413
Thr Pro Leu Gln Lys Ser Glu Val Val Lys Leu Val Arg Ser His Leu	
1035 1040 1045	
cag gtg atg acc ctt gct att ggt gat ggt gcc aat gat gtt agc atg	3461
Gln Val Met Thr Leu Ala Ile Gly Asp Gly Ala Asn Asp Val Ser Met	
1050 1055 1060	
ata caa gtg gca gac att ggg ata ggg gtc tca ggt caa gaa ggc atg	3509
Ile Gln Val Ala Asp Ile Gly Ile Gly Val Ser Gly Gln Glu Gly Met	
1065 1070 1075	
cag gct gtg atg gcc agt gac ttt gcc gtt tct cag ttc aaa cat ctc	3557
Gln Ala Val Met Ala Ser Asp Phe Ala Val Ser Gln Phe Lys His Leu	
1080 1085 1090 1095	

agc aag ctc ctt ctt gtc cat gga cac tgg tgt tat aca cgg ctt tcc	3605
Ser Lys Leu Leu Leu Val His Gly His Trp Cys Tyr Thr Arg Leu Ser	
1100 1105 1110	
aac atg att ctc tat ttt ttc tat aag aat gtg gcc tat gtg aac ctc	3653
Asn Met Ile Leu Tyr Phe Phe Tyr Lys Asn Val Ala Tyr Val Asn Leu	
1115 1120 1125	
ctt ttc tgg tac cag ttc ttt tgt gga ttt tca gga aca tcc atg act	3701
Leu Phe Trp Tyr Gln Phe Phe Cys Gly Phe Ser Gly Thr Ser Met Thr	
1130 1135 1140	
gat tac tgg gtt ttg atc ttc ttc aac ctc ctc ttc aca tct gcc cct	3749
Asp Tyr Trp Val Leu Ile Phe Phe Asn Leu Leu Phe Thr Ser Ala Pro	
1145 1150 1155	
cct gtc att tat ggt gtt ttg gag aaa gat gtg tct gca gag acc ctc	3797
Pro Val Ile Tyr Gly Val Leu Glu Lys Asp Val Ser Ala Glu Thr Leu	
1160 1165 1170 1175	
atg caa ctg cct gaa ctt tac aga agt ggt cag aaa tca gag gca tac	3845
Met Gln Leu Pro Glu Leu Tyr Arg Ser Gly Gln Lys Ser Glu Ala Tyr	
1180 1185 1190	
tta ccc cat acc ttc tgg atc acc tta ttg gat gct ttt tat caa agc	3893
Leu Pro His Thr Phe Trp Ile Thr Leu Leu Asp Ala Phe Tyr Gln Ser	
1195 1200 1205	
ctg gtc tgc ttc ttt gtg cct tat ttt acc tac cag ggc tca gat act	3941
Leu Val Cys Phe Phe Val Pro Tyr Phe Thr Tyr Gln Gly Ser Asp Thr	
1210 1215 1220	
gac atc ttt gca ttt gga aac ccc ctg aac aca gcc gct ctg ttc atc	3989
Asp Ile Phe Ala Phe Gly Asn Pro Leu Asn Thr Ala Ala Leu Phe Ile	
1225 1230 1235	
gtt ctc ctc cat ctg gtc att gaa agc aag agt ttg act tgg att cac	4037
Val Leu Leu His Leu Val Ile Glu Ser Lys Ser Leu Thr Trp Ile His	
1240 1245 1250 1255	
ttg ctg gtc atc att ggt agc atc ttg tct tat ttt tta ttt gcc ata	4085
Leu Leu Val Ile Ile Gly Ser Ile Leu Ser Tyr Phe Leu Phe Ala Ile	
1260 1265 1270	
gtt ttt gga gcc atg tgt gta act tgc aac cca cca tcc aac cct tac	4133
Val Phe Gly Ala Met Cys Val Thr Cys Asn Pro Pro Ser Asn Pro Tyr	
1275 1280 1285	
tgg att atg cag gag cac atg ctg gat cca gta ttc tac tta gtt tgt	4181
Trp Ile Met Gln Glu His Met Leu Asp Pro Val Phe Tyr Leu Val Cys	
1290 1295 1300	
atc ctc acg acg tcc att gct ctt ctg ccc agg ttt gta tac aga gtt	4229
Ile Leu Thr Thr Ser Ile Ala Leu Leu Pro Arg Phe Val Tyr Arg Val	
1305 1310 1315	
ctt cag gga tcc ctg ttt cca tct cca att ctg aga gct aag cac ttt	4277
Leu Gln Gly Ser Leu Phe Pro Ser Pro Ile Leu Arg Ala Lys His Phe	
1320 1325 1330 1335	

gac aga cta act cca gag gag agg act aaa gct ctc aag aag tgg aga 4325
 Asp Arg Leu Thr Pro Glu Glu Arg Thr Lys Ala Leu Lys Lys Trp Arg
 1340 1345 1350
 ggg gct gga aag atg aat caa gtg aca tca aag tat gct aac caa tca 4373
 Gly Ala Gly Lys Met Asn Gln Val Thr Ser Lys Tyr Ala Asn Gln Ser
 1355 1360 1365
 gct ggc aag tca gga aga aga ccc atg cct ggc cct tct gct gta ttt 4421
 Ala Gly Lys Ser Gly Arg Arg Pro Met Pro Gly Pro Ser Ala Val Phe
 1370 1375 1380
 gca atg aag tca gca act tcc tgt gct att gag caa gga aac tta tct 4469
 Ala Met Lys Ser Ala Thr Ser Cys Ala Ile Glu Gln Gly Asn Leu Ser
 1385 1390 1395
 ctg tgt gaa act gct tta gat caa ggc tac tct gaa act aag gcc ttt 4517
 Leu Cys Glu Thr Ala Leu Asp Gln Gly Tyr Ser Glu Thr Lys Ala Phe
 1400 1405 1410 1415
 gag atg gct gga ccc tcc aaa ggt aaa gaa agc tag ataccctcct 4563
 Glu Met Ala Gly Pro Ser Lys Gly Lys Glu Ser *
 1420 1425

tggagttgca agtattcttt caaggttgga agagggattt tgaagaggta tctctccaag 4623
 caagaatgac ttgtttttcc ataagggaca tgagcatttt actaggcttg gaagagctga 4683
 catgatgagc attattgtat gttgtatat acatttgtga tagagggcta gagtttgacc 4743
 tagagagagt ttaaggaagt gaaatatatta attcagaacc aaatgctttt gtaaaacttt 4803
 ttggattttg taaaagcatt ttcattctct tagaaattca agtattttca aggggagtc 4863
 tttgagatat atttatttta ctaggagatc ttatattcta gggaaatgct ttaaattggtc 4923
 aggtccaat cggaattttt ttaagaaaaa agtagttttt aatacattgg ttaggactca 4983
 gaggaataac ggaaaaaaca ttgtagatgg ttaatttaca gataaaatcc caagagcctt 5043
 ttaaacaca aggtacctaa aatagggtat aattatactg cttaaaatac agttagtgcc 5103
 tattaatagc tttttatttc ctatgggaag atgcttttgg tcttctggct gaagatgtag 5163
 gcatacctct cactcatttc aatgttttcc tgaggtggag ccttcattgg aaaggggaaa 5223
 gagggattct aggggttcat cagggaccag gaatgcattc ctctgtcagg ttccaatcaa 5283
 gagaagacct tttatgagat ctgcctctgt atagatgttg tcagttagga aactgaagcc 5343
 ataggtcagg cagacatcag ctccagcctgt ggcccatagg gtgatttctt gtattttaaa 5403
 actgacagta gcctgatcaa agtgatacaa tcaatttcaa aacaatcttc cagagaccac 5463
 ttgaaggttc atagttttta caataccctg agacttttca ggtgttgagg cctctaaaat 5523
 atgagatata aacagaaact aatacaagtt gttctctgga ggtttctatg aggttcttag 5583
 aaaaatttgg ttttaaaatc atttgaggac aggaatgtct atagcaagtt tactctatt 5643
 gcgaatcatg tatgctggct ttagttgtaa caaacgattt tattctaagt aaggccaggt 5703
 gctactataa aatcatatat tctgttgtaa gttcttttga agtcttatct ctatttatta 5763
 tatttgaaag ttgtcagcca ccagtcattc agaatttctt tcttgaatct ccatgctcat 5823
 atgcaatgtc tacatcaagg tcttcttaat gactattatt ctccagggttt agttttctac 5883
 cttctgccta ctatttttgt ctgacatttt tgtagccttc tgttattatt ggaaatagtc 5943
 tcttacataa gctgatttgc agaactttca aaatctcaca tagctaattg aagttgcttt 6003
 ctgctttctt atgactgttt ttataaataa actgtttcat aaataaaaaa aaaaaaaaaa 6063
 gggcggccgc 6073

<210> 17

<211> 1426

<212> PRT

<213> Homo sapiens

<400> 17

Met Thr Glu Ala Leu Gln Trp Ala Arg Tyr His Trp Arg Arg Leu Ile
 1 5 10 15
 Arg Gly Ala Thr Arg Asp Asp Asp Ser Gly Pro Tyr Asn Tyr Ser Ser
 20 25 30

Leu Leu Ala Cys Gly Arg Lys Ser Ser Gln Ile Pro Lys Leu Ser Gly
 35 40 45
 Arg His Arg Ile Val Val Pro His Ile Gln Pro Phe Lys Asp Glu Tyr
 50 55 60
 Glu Lys Phe Ser Gly Ala Tyr Val Asn Asn Arg Ile Arg Thr Thr Lys
 65 70 75 80
 Tyr Thr Leu Leu Asn Phe Val Pro Arg Asn Leu Phe Glu Gln Phe His
 85 90 95
 Arg Ala Ala Ser Leu Tyr Phe Leu Phe Leu Val Val Leu Asn Trp Val
 100 105 110
 Pro Leu Val Glu Ala Phe Gln Lys Glu Ile Thr Met Leu Pro Leu Val
 115 120 125
 Val Val Leu Thr Ile Ile Ala Ile Lys Asp Gly Leu Glu Asp Tyr Arg
 130 135 140
 Lys Tyr Lys Ile Asp Lys Glu Ile Asn Asn Leu Ile Thr Lys Val Tyr
 145 150 155 160
 Ser Arg Lys Glu Lys Lys Tyr Ile Asp Arg Cys Trp Lys Asp Val Thr
 165 170 175
 Val Gly Asp Phe Ile Arg Leu Ser Cys Asn Glu Val Ile Pro Ala Asp
 180 185 190
 Met Val Leu Leu Phe Ser Thr Asp Pro Asp Gly Ile Cys His Ile Glu
 195 200 205
 Thr Ser Gly Leu Asp Gly Glu Ser Asn Leu Lys Gln Arg Gln Val Val
 210 215 220
 Arg Gly Tyr Ala Glu Gln Asp Ser Glu Val Asp Pro Glu Lys Phe Ser
 225 230 235 240
 Ser Arg Ile Glu Cys Glu Ser Pro Asn Asn Asp Leu Ser Arg Phe Arg
 245 250 255
 Gly Phe Leu Glu His Ser Asn Lys Glu Arg Val Gly Leu Ser Lys Glu
 260 265 270
 Asn Leu Leu Leu Arg Gly Cys Thr Ile Arg Asn Thr Glu Ala Val Val
 275 280 285
 Gly Ile Val Val Tyr Ala Gly His Glu Thr Lys Ala Met Leu Asn Asn
 290 295 300
 Ser Gly Pro Arg Tyr Lys Arg Ser Lys Leu Glu Arg Arg Ala Asn Thr
 305 310 315 320
 Asp Val Leu Trp Cys Val Met Leu Leu Val Ile Met Cys Leu Thr Gly
 325 330 335
 Ala Val Gly His Gly Ile Trp Leu Ser Arg Tyr Glu Lys Met His Phe
 340 345 350
 Phe Asn Val Pro Glu Pro Asp Gly His Ile Ile Ser Pro Leu Leu Ala
 355 360 365
 Gly Phe Tyr Met Phe Trp Thr Met Ile Ile Leu Leu Gln Val Leu Ile
 370 375 380
 Pro Ile Ser Leu Tyr Val Ser Ile Glu Ile Val Lys Leu Gly Gln Ile
 385 390 395 400
 Tyr Phe Ile Gln Ser Asp Val Asp Phe Tyr Asn Glu Lys Met Asp Ser
 405 410 415
 Ile Val Gln Cys Arg Ala Leu Asn Ile Ala Glu Asp Leu Gly Gln Ile
 420 425 430
 Gln Tyr Leu Phe Ser Asp Lys Thr Gly Thr Leu Thr Glu Asn Lys Met
 435 440 445
 Val Phe Arg Arg Cys Ser Val Ala Gly Phe Asp Tyr Cys His Glu Glu
 450 455 460
 Asn Ala Arg Arg Leu Glu Ser Tyr Gln Glu Ala Val Ser Glu Asp Glu
 465 470 475 480
 Asp Phe Ile Asp Thr Val Ser Gly Ser Leu Ser Asn Met Ala Lys Pro
 485 490 495
 Arg Ala Pro Ser Cys Arg Thr Val His Asn Gly Pro Leu Gly Asn Lys
 500 505 510

Pro	Ser	Asn	His	Leu	Ala	Gly	Ser	Ser	Phe	Thr	Leu	Gly	Ser	Gly	Glu
		515					520					525			
Gly	Ala	Ser	Glu	Val	Pro	His	Ser	Arg	Gln	Ala	Ala	Phe	Ser	Ser	Pro
530						535					540				
Ile	Glu	Thr	Asp	Val	Val	Pro	Asp	Thr	Arg	Leu	Leu	Asp	Lys	Phe	Ser
545					550					555					560
Gln	Ile	Thr	Pro	Arg	Leu	Phe	Met	Pro	Leu	Asp	Glu	Thr	Ile	Gln	Asn
				565					570					575	
Pro	Pro	Met	Glu	Thr	Leu	Tyr	Ile	Ile	Asp	Phe	Phe	Ile	Ala	Leu	Ala
				580				585					590		
Ile	Cys	Asn	Thr	Val	Val	Val	Ser	Ala	Pro	Asn	Gln	Pro	Arg	Gln	Lys
		595					600					605			
Ile	Arg	His	Pro	Ser	Leu	Gly	Gly	Leu	Pro	Ile	Lys	Ser	Leu	Glu	Glu
	610					615					620				
Ile	Lys	Ser	Leu	Phe	Gln	Arg	Trp	Ser	Val	Arg	Arg	Ser	Ser	Ser	Pro
625					630					635					640
Ser	Leu	Asn	Ser	Gly	Lys	Glu	Pro	Ser	Ser	Gly	Val	Pro	Asn	Ala	Phe
				645					650					655	
Val	Ser	Arg	Leu	Pro	Leu	Phe	Ser	Arg	Met	Lys	Pro	Ala	Ser	Pro	Val
			660					665					670		
Glu	Glu	Glu	Val	Ser	Gln	Val	Cys	Glu	Ser	Pro	Gln	Cys	Ser	Ser	Ser
			675				680					685			
Ser	Ala	Cys	Cys	Thr	Glu	Thr	Glu	Lys	Gln	His	Gly	Asp	Ala	Gly	Leu
	690					695					700				
Leu	Asn	Gly	Lys	Ala	Glu	Ser	Leu	Pro	Gly	Gln	Pro	Leu	Ala	Cys	Asn
705					710					715					720
Leu	Cys	Tyr	Glu	Ala	Glu	Ser	Pro	Asp	Glu	Ala	Ala	Leu	Val	Tyr	Ala
				725					730					735	
Ala	Arg	Ala	Tyr	Gln	Cys	Thr	Leu	Arg	Ser	Arg	Thr	Pro	Glu	Gln	Val
			740					745					750		
Met	Val	Asp	Phe	Ala	Ala	Leu	Gly	Pro	Leu	Thr	Phe	Gln	Leu	Leu	His
		755					760					765			
Ile	Leu	Pro	Phe	Asp	Ser	Val	Arg	Lys	Arg	Met	Ser	Val	Val	Val	Arg
	770					775					780				
His	Pro	Leu	Ser	Asn	Gln	Val	Val	Val	Tyr	Thr	Lys	Gly	Ala	Asp	Ser
785					790					795					800
Val	Ile	Met	Glu	Leu	Ser	Val	Ala	Ser	Pro	Asp	Gly	Ala	Ser	Leu	
				805					810					815	
Glu	Lys	Gln	Gln	Met	Ile	Val	Arg	Glu	Lys	Thr	Gln	Lys	His	Leu	Asp
			820					825					830		
Asp	Tyr	Ala	Lys	Gln	Gly	Leu	Arg	Thr	Leu	Cys	Ile	Ala	Lys	Lys	Val
		835					840					845			
Met	Ser	Asp	Thr	Glu	Tyr	Ala	Glu	Trp	Leu	Arg	Asn	His	Phe	Leu	Ala
		850				855					860				
Glu	Thr	Ser	Ile	Asp	Asn	Arg	Glu	Glu	Leu	Leu	Leu	Glu	Ser	Ala	Met
865					870				875						880
Arg	Leu	Glu	Asn	Lys	Leu	Thr	Leu	Leu	Gly	Ala	Thr	Gly	Ile	Glu	Asp
				885					890					895	
Arg	Leu	Gln	Glu	Gly	Val	Pro	Glu	Ser	Ile	Glu	Ala	Leu	His	Lys	Ala
			900					905					910		
Gly	Ile	Lys	Ile	Trp	Met	Leu	Thr	Gly	Asp	Lys	Gln	Glu	Thr	Ala	Val
		915					920					925			
Asn	Ile	Ala	Tyr	Ala	Cys	Lys	Leu	Leu	Glu	Pro	Asp	Asp	Lys	Leu	Phe
		930				935					940				
Ile	Leu	Asn	Thr	Gln	Ser	Lys	Asp	Ala	Cys	Gly	Met	Leu	Met	Ser	Thr
945					950					955					960
Ile	Leu	Lys	Glu	Leu	Gln	Lys	Lys	Thr	Gln	Ala	Leu	Pro	Glu	Gln	Val
				965					970					975	
Ser	Leu	Ser	Glu	Asp	Leu	Leu	Gln	Pro	Pro	Val	Pro	Arg	Asp	Ser	Gly
			980					985					990		

Leu Arg Ala Gly Leu Ile Ile Thr Gly Lys Thr Leu Glu Phe Ala Leu
 995 1000 1005
 Gln Glu Ser Leu Gln Lys Gln Phe Leu Glu Leu Thr Ser Trp Cys Gln
 1010 1015 1020
 Ala Val Val Cys Cys Arg Ala Thr Pro Leu Gln Lys Ser Glu Val Val
 1025 1030 1035 1040
 Lys Leu Val Arg Ser His Leu Gln Val Met Thr Leu Ala Ile Gly Asp
 1045 1050 1055
 Gly Ala Asn Asp Val Ser Met Ile Gln Val Ala Asp Ile Gly Ile Gly
 1060 1065 1070
 Val Ser Gly Gln Glu Gly Met Gln Ala Val Met Ala Ser Asp Phe Ala
 1075 1080 1085
 Val Ser Gln Phe Lys His Leu Ser Lys Leu Leu Leu Val His Gly His
 1090 1095 1100
 Trp Cys Tyr Thr Arg Leu Ser Asn Met Ile Leu Tyr Phe Phe Tyr Lys
 1105 1110 1115 1120
 Asn Val Ala Tyr Val Asn Leu Leu Phe Trp Tyr Gln Phe Phe Cys Gly
 1125 1130 1135
 Phe Ser Gly Thr Ser Met Thr Asp Tyr Trp Val Leu Ile Phe Phe Asn
 1140 1145 1150
 Leu Leu Phe Thr Ser Ala Pro Pro Val Ile Tyr Gly Val Leu Glu Lys
 1155 1160 1165
 Asp Val Ser Ala Glu Thr Leu Met Gln Leu Pro Glu Leu Tyr Arg Ser
 1170 1175 1180
 Gly Gln Lys Ser Glu Ala Tyr Leu Pro His Thr Phe Trp Ile Thr Leu
 1185 1190 1195 1200
 Leu Asp Ala Phe Tyr Gln Ser Leu Val Cys Phe Phe Val Pro Tyr Phe
 1205 1210 1215
 Thr Tyr Gln Gly Ser Asp Thr Asp Ile Phe Ala Phe Gly Asn Pro Leu
 1220 1225 1230
 Asn Thr Ala Ala Leu Phe Ile Val Leu Leu His Leu Val Ile Glu Ser
 1235 1240 1245
 Lys Ser Leu Thr Trp Ile His Leu Leu Val Ile Ile Gly Ser Ile Leu
 1250 1255 1260
 Ser Tyr Phe Leu Phe Ala Ile Val Phe Gly Ala Met Cys Val Thr Cys
 1265 1270 1275 1280
 Asn Pro Pro Ser Asn Pro Tyr Trp Ile Met Gln Glu His Met Leu Asp
 1285 1290 1295
 Pro Val Phe Tyr Leu Val Cys Ile Leu Thr Thr Ser Ile Ala Leu Leu
 1300 1305 1310
 Pro Arg Phe Val Tyr Arg Val Leu Gln Gly Ser Leu Phe Pro Ser Pro
 1315 1320 1325
 Ile Leu Arg Ala Lys His Phe Asp Arg Leu Thr Pro Glu Arg Thr
 1330 1335 1340
 Lys Ala Leu Lys Lys Trp Arg Gly Ala Gly Lys Met Asn Gln Val Thr
 1345 1350 1355 1360
 Ser Lys Tyr Ala Asn Gln Ser Ala Gly Lys Ser Gly Arg Arg Pro Met
 1365 1370 1375
 Pro Gly Pro Ser Ala Val Phe Ala Met Lys Ser Ala Thr Ser Cys Ala
 1380 1385 1390
 Ile Glu Gln Gly Asn Leu Ser Leu Cys Glu Thr Ala Leu Asp Gln Gly
 1395 1400 1405
 Tyr Ser Glu Thr Lys Ala Phe Glu Met Ala Gly Pro Ser Lys Gly Lys
 1410 1415 1420
 Glu Ser
 1425

<210> 18

<211> 4281

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(4281)

<400> 18

atg act gag gct ctc caa tgg gcc aga tat cac tgg cga cgg ctg atc 48
 Met Thr Glu Ala Leu Gln Trp Ala Arg Tyr His Trp Arg Arg Leu Ile
 1 5 10 15

aga ggt gca acc agg gat gat gat tca ggg cca tac aac tat tcc tcg 96
 Arg Gly Ala Thr Arg Asp Asp Asp Ser Gly Pro Tyr Asn Tyr Ser Ser
 20 25 30

ttg ctc gcc tgt ggg cgc aag tcc tct cag atc cct aaa ctg tca gga 144
 Leu Leu Ala Cys Gly Arg Lys Ser Ser Gln Ile Pro Lys Leu Ser Gly
 35 40 45

agg cac cgg att gtt gtt ccc cac atc cag ccc ttc aag gat gag tat 192
 Arg His Arg Ile Val Val Pro His Ile Gln Pro Phe Lys Asp Glu Tyr
 50 55 60

gag aag ttc tcc gga gcc tat gtg aac aat cga ata cga aca aca aag 240
 Glu Lys Phe Ser Gly Ala Tyr Val Asn Asn Arg Ile Arg Thr Thr Lys
 65 70 75 80

tac aca ctt ctg aat ttt gtg cca aga aat tta ttt gaa caa ttt cac 288
 Tyr Thr Leu Leu Asn Phe Val Pro Arg Asn Leu Phe Glu Gln Phe His
 85 90 95

aga gct gcc agt tta tat ttc ctg ttc cta gtt gtc ctg aac tgg gta 336
 Arg Ala Ala Ser Leu Tyr Phe Leu Phe Leu Val Val Leu Asn Trp Val
 100 105 110

cct ttg gta gaa gcc ttc caa aag gaa atc acc atg ttg cct ctg gtg 384
 Pro Leu Val Glu Ala Phe Gln Lys Glu Ile Thr Met Leu Pro Leu Val
 115 120 125

gtg gtc ctt aca att atc gca att aaa gat ggc ctg gaa gat tat cgg 432
 Val Val Leu Thr Ile Ile Ala Ile Lys Asp Gly Leu Glu Asp Tyr Arg
 130 135 140

aaa tac aaa att gac aaa cag atc aat aat tta ata act aaa gtt tat 480
 Lys Tyr Lys Ile Asp Lys Gln Ile Asn Asn Leu Ile Thr Lys Val Tyr
 145 150 155 160

agt agg aaa gag aaa aaa tac att gac cga tgc tgg aaa gac gtt act 528
 Ser Arg Lys Glu Lys Lys Tyr Ile Asp Arg Cys Trp Lys Asp Val Thr
 165 170 175

gtt ggg gac ttt att cgc ctc tcc tgc aat gag gtc atc cct gca gac 576
 Val Gly Asp Phe Ile Arg Leu Ser Cys Asn Glu Val Ile Pro Ala Asp
 180 185 190

atg gta cta ctc ttt tcc act gat cca gat gga atc tgt cac att gag 624
 Met Val Leu Leu Phe Ser Thr Asp Pro Asp Gly Ile Cys His Ile Glu
 195 200 205

act tct ggt ctt gat gga gag agc aat tta aaa cag agg cag gtg gtt	672
Thr Ser Gly Leu Asp Gly Glu Ser Asn Leu Lys Gln Arg Gln Val Val	
210 215 220	
cgg gga tat gca gaa cag gac tct gaa gtt gat cct gag aag ttt tcc	720
Arg Gly Tyr Ala Glu Gln Asp Ser Glu Val Asp Pro Glu Lys Phe Ser	
225 230 235 240	
agt agg ata gaa tgt gaa agc cca aac aat gac ctc agc aga ttc cga	768
Ser Arg Ile Glu Cys Glu Ser Pro Asn Asn Asp Leu Ser Arg Phe Arg	
245 250 255	
ggc ttc cta gaa cat tcc aac aaa gaa cgc gtg ggt ctc agt aaa gaa	816
Gly Phe Leu Glu His Ser Asn Lys Glu Arg Val Gly Leu Ser Lys Glu	
260 265 270	
aat ttg ttg ctt aga gga tgc acc att aga aac aca gag gct gtt gtg	864
Asn Leu Leu Leu Arg Gly Cys Thr Ile Arg Asn Thr Glu Ala Val Val	
275 280 285	
ggc att gtg gtt tat gca ggc cat gaa acc aaa gca atg ctg aac aac	912
Gly Ile Val Val Tyr Ala Gly His Glu Thr Lys Ala Met Leu Asn Asn	
290 295 300	
agt ggg cca cgg tat aag cgc agc aaa tta gaa aga aga gca aac aca	960
Ser Gly Pro Arg Tyr Lys Arg Ser Lys Leu Glu Arg Arg Ala Asn Thr	
305 310 315 320	
gat gtc ctc tgg tgt gtc atg ctt ctg gtc ata atg tgc tta act ggc	1008
Asp Val Leu Trp Cys Val Met Leu Leu Val Ile Met Cys Leu Thr Gly	
325 330 335	
gca gta ggt cat gga atc tgg ctg agc agg tat gaa aag atg cat ttt	1056
Ala Val Gly His Gly Ile Trp Leu Ser Arg Tyr Glu Lys Met His Phe	
340 345 350	
ttc aat gtt ccc gag cct gat gga cat atc ata tca cca ctg ttg gca	1104
Phe Asn Val Pro Glu Pro Asp Gly His Ile Ile Ser Pro Leu Leu Ala	
355 360 365	
gga ttt tat atg ttt tgg acc atg atc att ttg tta cag gtc ttg att	1152
Gly Phe Tyr Met Phe Trp Thr Met Ile Ile Leu Leu Gln Val Leu Ile	
370 375 380	
cct att tct ctc tat gtt tcc atc gaa att gtg aag ctt gga caa ata	1200
Pro Ile Ser Leu Tyr Val Ser Ile Glu Ile Val Lys Leu Gly Gln Ile	
385 390 395 400	
tat ttc att caa agt gat gtg gat ttc tac aat gaa aaa atg gat tct	1248
Tyr Phe Ile Gln Ser Asp Val Asp Phe Tyr Asn Glu Lys Met Asp Ser	
405 410 415	
att gtt cag tgc cga gcc ctg aac atc gcc gag gat ctg gga cag att	1296
Ile Val Gln Cys Arg Ala Leu Asn Ile Ala Glu Asp Leu Gly Gln Ile	
420 425 430	
cag tac ctc ttt tcc gat aag aca gga acc ctc act gag aat aag atg	1344
Gln Tyr Leu Phe Ser Asp Lys Thr Gly Thr Leu Thr Glu Asn Lys Met	
435 440 445	

gtt ttt cga aga tgt agt gtg gca gga ttt gat tac tgc cat gaa gaa	1392
Val Phe Arg Arg Cys Ser Val Ala Gly Phe Asp Tyr Cys His Glu Glu	
450 455 460	
aat gcc agg agg ttg gag tcc tat cag gaa gct gtc tct gaa gat gaa	1440
Asn Ala Arg Arg Leu Glu Ser Tyr Gln Glu Ala Val Ser Glu Asp Glu	
465 470 475 480	
gat ttt ata gac aca gtc agt ggt tcc ctc agc aat atg gca aaa ccg	1488
Asp Phe Ile Asp Thr Val Ser Gly Ser Leu Ser Asn Met Ala Lys Pro	
485 490 495	
aga gcc ccc agc tgc agg aca gtt cat aat ggg cct ttg gga aat aag	1536
Arg Ala Pro Ser Cys Arg Thr Val His Asn Gly Pro Leu Gly Asn Lys	
500 505 510	
ccc tca aat cat ctt gct ggg agc tct ttt act cta gga agt gga gaa	1584
Pro Ser Asn His Leu Ala Gly Ser Ser Phe Thr Leu Gly Ser Gly Glu	
515 520 525	
gga gcc agt gaa gtg cct cat tcc aga cag gct gct ttc agt agc ccc	1632
Gly Ala Ser Glu Val Pro His Ser Arg Gln Ala Ala Phe Ser Ser Pro	
530 535 540	
att gaa aca gac gtg gta cca gac acc agg ctt tta gac aaa ttt agt	1680
Ile Glu Thr Asp Val Val Pro Asp Thr Arg Leu Leu Asp Lys Phe Ser	
545 550 555 560	
cag att aca cct cgg ctc ttt atg cca cta gat gag acc atc caa aat	1728
Gln Ile Thr Pro Arg Leu Phe Met Pro Leu Asp Glu Thr Ile Gln Asn	
565 570 575	
cca cca atg gaa act ttg tac att atc gac ttt ttc att gca ttg gca	1776
Pro Pro Met Glu Thr Leu Tyr Ile Ile Asp Phe Phe Ile Ala Leu Ala	
580 585 590	
att tgc aac aca gta gtg gtt tct gct cct aac caa ccc cga caa aag	1824
Ile Cys Asn Thr Val Val Val Ser Ala Pro Asn Gln Pro Arg Gln Lys	
595 600 605	
atc aga cac cct tca ctg ggg ggg ttg ccc att aag tct ttg gaa gag	1872
Ile Arg His Pro Ser Leu Gly Gly Leu Pro Ile Lys Ser Leu Glu Glu	
610 615 620	
att aaa agt ctt ttc cag aga tgg tct gtc cga aga tca agt tct cca	1920
Ile Lys Ser Leu Phe Gln Arg Trp Ser Val Arg Arg Ser Ser Ser Pro	
625 630 635 640	
tcg ctt aac agt ggg aaa gag cca tct tct gga gtt cca aac gcc ttt	1968
Ser Leu Asn Ser Gly Lys Glu Pro Ser Ser Gly Val Pro Asn Ala Phe	
645 650 655	
gtg agc aga ctc cct ctc ttt agt cga atg aaa cca gct tca cct gtg	2016
Val Ser Arg Leu Pro Leu Phe Ser Arg Met Lys Pro Ala Ser Pro Val	
660 665 670	
gag gaa gag gtc tcc cag gtg tgt gag agc ccc cag tgc tcc agt agc	2064
Glu Glu Glu Val Ser Gln Val Cys Glu Ser Pro Gln Cys Ser Ser Ser	
675 680 685	

tca gct tgc tgc aca gaa aca gag aaa caa cac ggt gat gca ggc ctc	2112
Ser Ala Cys Cys Thr Glu Thr Glu Lys Gln His Gly Asp Ala Gly Leu	
690 695 700	
ctg aat ggc aag gca gag tcc ctc cct gga cag cca ttg gcc tgc aac	2160
Leu Asn Gly Lys Ala Glu Ser Leu Pro Gly Gln Pro Leu Ala Cys Asn	
705 710 715 720	
ctg tgt tat gag gcc gag agc cca gac gaa gcg gcc tta gtg tat gcc	2208
Leu Cys Tyr Glu Ala Glu Ser Pro Asp Glu Ala Ala Leu Val Tyr Ala	
725 730 735	
gcc agg gct tac caa tgc act tta cgg tct cgg aca cca gag cag gtc	2256
Ala Arg Ala Tyr Gln Cys Thr Leu Arg Ser Arg Thr Pro Glu Gln Val	
740 745 750	
atg gtg gac ttt gct gct ttg gga cca tta aca ttt caa ctc cta cac	2304
Met Val Asp Phe Ala Ala Leu Gly Pro Leu Thr Phe Gln Leu Leu His	
755 760 765	
atc ctg ccc ttt gac tca gta aga aaa aga atg tct gtt gtg gtc cga	2352
Ile Leu Pro Phe Asp Ser Val Arg Lys Arg Met Ser Val Val Val Arg	
770 775 780	
cac cct ctt tcc aat caa gtt gtg gtg tat acg aaa ggc gct gat tct	2400
His Pro Leu Ser Asn Gln Val Val Val Tyr Thr Lys Gly Ala Asp Ser	
785 790 795 800	
gtg atc atg gag tta ctg tcg gtg gct tcc cca gat gga gca agt ctg	2448
Val Ile Met Glu Leu Leu Ser Val Ala Ser Pro Asp Gly Ala Ser Leu	
805 810 815	
gag aaa caa cag atg ata gta agg gag aaa acc cag aag cac ttg gat	2496
Glu Lys Gln Gln Met Ile Val Arg Glu Lys Thr Gln Lys His Leu Asp	
820 825 830	
gac tat gcc aaa caa ggc ctt cgt act tta tgt ata gca aag aag gtc	2544
Asp Tyr Ala Lys Gln Gly Leu Arg Thr Leu Cys Ile Ala Lys Lys Val	
835 840 845	
atg agt gac act gaa tat gca gag tgg ctg agg aat cat ttt tta gct	2592
Met Ser Asp Thr Glu Tyr Ala Glu Trp Leu Arg Asn His Phe Leu Ala	
850 855 860	
gaa acc agc att gac aac agg gaa gaa tta cta ctt gaa tct gcc atg	2640
Glu Thr Ser Ile Asp Asn Arg Glu Glu Leu Leu Leu Glu Ser Ala Met	
865 870 875 880	
agg ttg gag aac aaa ctt aca tta ctt ggt gct act ggc att gaa gac	2688
Arg Leu Glu Asn Lys Leu Thr Leu Leu Gly Ala Thr Gly Ile Glu Asp	
885 890 895	
cgt ctg cag gag gga gtc cct gaa tct ata gaa gct ctt cac aaa gcg	2736
Arg Leu Gln Glu Gly Val Pro Glu Ser Ile Glu Ala Leu His Lys Ala	
900 905 910	
ggc atc aag atc tgg atg ctg aca ggg gac aag cag gag aca gct gtc	2784
Gly Ile Lys Ile Trp Met Leu Thr Gly Asp Lys Gln Glu Thr Ala Val	
915 920 925	

aac ata gct tat gca tgc aaa cta ctg gag cca gat gac aag ctt ttt	2832
Asn Ile Ala Tyr Ala Cys Lys Leu Leu Glu Pro Asp Asp Lys Leu Phe	
930 935 940	
atc ctc aat acc caa agt aaa gat gcc tgt ggg atg ctg atg agc aca	2880
Ile Leu Asn Thr Gln Ser Lys Asp Ala Cys Gly Met Leu Met Ser Thr	
945 950 955 960	
att ttg aaa gaa ctt cag aag aaa act caa gcc ctg cca gag caa gtg	2928
Ile Leu Lys Glu Leu Gln Lys Lys Thr Gln Ala Leu Pro Glu Gln Val	
965 970 975	
tca tta agt gaa gat tta ctt cag cct cct gtc ccc cgg gac tca ggg	2976
Ser Leu Ser Glu Asp Leu Leu Gln Pro Pro Val Pro Arg Asp Ser Gly	
980 985 990	
tta cga gct gga ctc att atc act ggg aag acc ctg gag ttt gcc ctg	3024
Leu Arg Ala Gly Leu Ile Ile Thr Gly Lys Thr Leu Glu Phe Ala Leu	
995 1000 1005	
caa gaa agt ctg caa aag cag ttc ctg gaa ctg aca tct tgg tgt caa	3072
Gln Glu Ser Leu Gln Lys Gln Phe Leu Glu Leu Thr Ser Trp Cys Gln	
1010 1015 1020	
gct gtg gtc tgc tgc cga gcc aca ccg ctg cag aaa agt gaa gtg gtg	3120
Ala Val Val Cys Cys Arg Ala Thr Pro Leu Gln Lys Ser Glu Val Val	
1025 1030 1035 1040	
aaa ttg gtc cgc agc cat ctc cag gtg atg acc ctt gct att ggt gat	3168
Lys Leu Val Arg Ser His Leu Gln Val Met Thr Leu Ala Ile Gly Asp	
1045 1050 1055	
ggt gcc aat gat gtt agc atg ata caa gtg gca gac att ggg ata ggg	3216
Gly Ala Asn Asp Val Ser Met Ile Gln Val Ala Asp Ile Gly Ile Gly	
1060 1065 1070	
gtc tca ggt caa gaa ggc atg cag gct gtg atg gcc agt gac ttt gcc	3264
Val Ser Gly Gln Glu Gly Met Gln Ala Val Met Ala Ser Asp Phe Ala	
1075 1080 1085	
gtt tct cag ttc aaa cat ctc agc aag ctc ctt ctt gtc cat gga cac	3312
Val Ser Gln Phe Lys His Leu Ser Lys Leu Leu Leu Val His Gly His	
1090 1095 1100	
tgg tgt tat aca cgg ctt tcc aac atg att ctc tat ttt ttc tat aag	3360
Trp Cys Tyr Thr Arg Leu Ser Asn Met Ile Leu Tyr Phe Phe Tyr Lys	
1105 1110 1115 1120	
aat gtg gcc tat gtg aac ctc ctt ttc tgg tac cag ttc ttt tgt gga	3408
Asn Val Ala Tyr Val Asn Leu Leu Phe Trp Tyr Gln Phe Phe Cys Gly	
1125 1130 1135	
ttt tca gga aca tcc atg act gat tac tgg gtt ttg atc ttc ttc aac	3456
Phe Ser Gly Thr Ser Met Thr Asp Tyr Trp Val Leu Ile Phe Phe Asn	
1140 1145 1150	
ctc ctc ttc aca tct gcc cct cct gtc att tat ggt gtt ttg gag aaa	3504
Leu Leu Phe Thr Ser Ala Pro Pro Val Ile Tyr Gly Val Leu Glu Lys	
1155 1160 1165	

gat	gtg	tct	gca	gag	acc	ctc	atg	caa	ctg	cct	gaa	ctt	tac	aga	agt	3552
Asp	Val	Ser	Ala	Glu	Thr	Leu	Met	Gln	Leu	Pro	Glu	Leu	Tyr	Arg	Ser	
	1170					1175					1180					
ggt	cag	aaa	tca	gag	gca	tac	tta	ccc	cat	acc	ttc	tgg	atc	acc	tta	3600
Gly	Gln	Lys	Ser	Glu	Ala	Tyr	Leu	Pro	His	Thr	Phe	Trp	Ile	Thr	Leu	
	1185					1190				1195					1200	
ttg	gat	gct	ttt	tat	caa	agc	ctg	gtc	tgc	ttc	ttt	gtg	cct	tat	ttt	3648
Leu	Asp	Ala	Phe	Tyr	Gln	Ser	Leu	Val	Cys	Phe	Phe	Val	Pro	Tyr	Phe	
				1205					1210						1215	
acc	tac	cag	ggc	tca	gat	act	gac	atc	ttt	gca	ttt	gga	aac	ccc	ctg	3696
Thr	Tyr	Gln	Gly	Ser	Asp	Thr	Asp	Ile	Phe	Ala	Phe	Gly	Asn	Pro	Leu	
			1220					1225						1230		
aac	aca	gcc	gct	ctg	ttc	atc	ggt	ctc	ctc	cat	ctg	gtc	att	gaa	agc	3744
Asn	Thr	Ala	Ala	Leu	Phe	Ile	Val	Leu	Leu	His	Leu	Val	Ile	Glu	Ser	
		1235					1240					1245				
aag	agt	ttg	act	tgg	att	cac	ttg	ctg	gtc	atc	att	ggt	agg	atc	ttg	3792
Lys	Ser	Leu	Thr	Trp	Ile	His	Leu	Leu	Val	Ile	Ile	Gly	Ser	Ile	Leu	
	1250					1255					1260					
tct	tat	ttt	tta	ttt	gcc	ata	ggt	ttt	gga	gcc	atg	tgt	gta	act	tgc	3840
Ser	Tyr	Phe	Leu	Phe	Ala	Ile	Val	Phe	Gly	Ala	Met	Cys	Val	Thr	Cys	
	1265					1270				1275					1280	
aac	cca	cca	tcc	aac	cct	tac	tgg	att	atg	cag	gag	cac	atg	ctg	gat	3888
Asn	Pro	Pro	Ser	Asn	Pro	Tyr	Trp	Ile	Met	Gln	Glu	His	Met	Leu	Asp	
				1285					1290						1295	
cca	gta	ttc	tac	tta	ggt	tgt	atc	ctc	acg	acg	tcc	att	gct	ctt	ctg	3936
Pro	Val	Phe	Tyr	Leu	Val	Cys	Ile	Leu	Thr	Thr	Ser	Ile	Ala	Leu	Leu	
			1300					1305					1310			
ccc	agg	ttt	gta	tac	aga	ggt	ctt	cag	gga	tcc	ctg	ttt	cca	tct	cca	3984
Pro	Arg	Phe	Val	Tyr	Arg	Val	Leu	Gln	Gly	Ser	Leu	Phe	Pro	Ser	Pro	
		1315					1320					1325				
att	ctg	aga	gct	aag	cac	ttt	gac	aga	cta	act	cca	gag	gag	agg	act	4032
Ile	Leu	Arg	Ala	Lys	His	Phe	Asp	Arg	Leu	Thr	Pro	Glu	Glu	Arg	Thr	
	1330					1335					1340					
aaa	gct	ctc	aag	aag	tgg	aga	ggg	gct	gga	aag	atg	aat	caa	gtg	aca	4080
Lys	Ala	Leu	Lys	Lys	Trp	Arg	Gly	Ala	Gly	Lys	Met	Asn	Gln	Val	Thr	
	1345					1350				1355					1360	
tca	aag	tat	gct	aac	caa	tca	gct	ggc	aag	tca	gga	aga	aga	ccc	atg	4128
Ser	Lys	Tyr	Ala	Asn	Gln	Ser	Ala	Gly	Lys	Ser	Gly	Arg	Arg	Pro	Met	
				1365				1370						1375		
cct	ggc	cct	tct	gct	gta	ttt	gca	atg	aag	tca	gca	act	tcc	tgt	gct	4176
Pro	Gly	Pro	Ser	Ala	Val	Phe	Ala	Met	Lys	Ser	Ala	Thr	Ser	Cys	Ala	
			1380					1385					1390			
att	gag	caa	gga	aac	tta	tct	ctg	tgt	gaa	act	gct	tta	gat	caa	ggc	4224
Ile	Glu	Gln	Gly	Asn	Leu	Ser	Leu	Cys	Glu	Thr	Ala	Leu	Asp	Gln	Gly	
		1395					1400					1405				

tac tct gaa act aag gcc ttt gag atg gct gga ccc tcc aaa ggt aaa 4272
 Tyr Ser Glu Thr Lys Ala Phe Glu Met Ala Gly Pro Ser Lys Gly Lys
 1410 1415 1420

gaa agc tag 4281
 Glu Ser *
 1425

<210> 19
 <211> 7221
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (167) ... (3700)

<400> 19
 gccgcgggat gggaacgcgg cgcggggagt gaggcagtgg cggcggcggc ggtaagcgga 60
 acttcggccc gaggggctcg cccgctcccg cctctgtctt gtcggcctcc acctgcagcc 120
 ccgcggcccc cgcgccccgc gggaccgcga cggcgacgac ggggga atg tgg cgc 175
 Met Trp Arg
 1

tgg atc cgg cag cag ctg ggt ttt gac cca cca cat cag agt gac aca 223
 Trp Ile Arg Gln Gln Leu Gly Phe Asp Pro Pro His Gln Ser Asp Thr
 5 10 15

aga acc atc tac gta gcc cac agg ttt cct cag aat ggc ctt tac aca 271
 Arg Thr Ile Tyr Val Ala His Arg Phe Pro Gln Asn Gly Leu Tyr Thr
 20 25 30 35

cct cag aaa ttt ata gat aac agg atc att tca tct aag tac act gtg 319
 Pro Gln Lys Phe Ile Asp Asn Arg Ile Ile Ser Ser Lys Tyr Thr Val
 40 45 50

tgg aat ttt gtt cca aaa aat tta ttt gaa cag ttc aga aga gtg gca 367
 Trp Asn Phe Val Pro Lys Asn Leu Phe Glu Gln Phe Arg Arg Val Ala
 55 60 65

aac ttt tat ttt ctt att ata ttt ttg gtt cag ctt atg att gat aca 415
 Asn Phe Tyr Phe Leu Ile Ile Phe Leu Val Gln Leu Met Ile Asp Thr
 70 75 80

cct acc agt cca gtt acc agt gga ctt cca tta ttc ttt gtg ata aca 463
 Pro Thr Ser Pro Val Thr Ser Gly Leu Pro Leu Phe Phe Val Ile Thr
 85 90 95

gta act gcc ata aag cag gga tat gaa gat tgg tta cgg cat aac tca 511
 Val Thr Ala Ile Lys Gln Gly Tyr Glu Asp Trp Leu Arg His Asn Ser
 100 105 110 115

gat aat gaa gta aat gga gct cct gtt tat gtt gtt cga agt ggt ggc 559
 Asp Asn Glu Val Asn Gly Ala Pro Val Tyr Val Val Arg Ser Gly Gly
 120 125 130

ctt gta aaa act aga tca aaa aac att cgg gtg ggt gat att gtt cga 607
 Leu Val Lys Thr Arg Ser Lys Asn Ile Arg Val Gly Asp Ile Val Arg
 135 140 145

ata gcc aaa gat gaa att ttt cct gca gac ttg gtg ctt ctg tcc tca	655
Ile Ala Lys Asp Glu Ile Phe Pro Ala Asp Leu Val Leu Leu Ser Ser	
150 155 160	
gat cga ctg gat ggt tcc tgt cac gtt aca act gct agt ttg gac gga	703
Asp Arg Leu Asp Gly Ser Cys His Val Thr Thr Ala Ser Leu Asp Gly	
165 170 175	
gaa act aac ctg aag aca cat gtg gca gtt cca gaa aca gca tta tta	751
Glu Thr Asn Leu Lys Thr His Val Ala Val Pro Glu Thr Ala Leu Leu	
180 185 190 195	
caa aca gtt gcc aat ttg gac act cta gta gct gta ata gaa tgc cag	799
Gln Thr Val Ala Asn Leu Asp Thr Leu Val Ala Val Ile Glu Cys Gln	
200 205 210	
caa cca gaa gca gac tta tac aga ttc atg gga cga atg atc ata acc	847
Gln Pro Glu Ala Asp Leu Tyr Arg Phe Met Gly Arg Met Ile Ile Thr	
215 220 225	
caa caa atg gaa gaa att gta aga cct ctg ggg ccg gag agt ctc ctg	895
Gln Gln Met Glu Glu Ile Val Arg Pro Leu Gly Pro Glu Ser Leu Leu	
230 235 240	
ctt cgt gga gcc aga tta aaa aac aca aaa gaa att ttt ggt gtt gcg	943
Leu Arg Gly Ala Arg Leu Lys Asn Thr Lys Glu Ile Phe Gly Val Ala	
245 250 255	
gta tac act gga atg gaa act aag atg gca tta aat tac aag agc aaa	991
Val Tyr Thr Gly Met Glu Thr Lys Met Ala Leu Asn Tyr Lys Ser Lys	
260 265 270 275	
tca cag aaa cga tct gca gta gaa aag tca atg aat aca ttt ttg ata	1039
Ser Gln Lys Arg Ser Ala Val Glu Lys Ser Met Asn Thr Phe Leu Ile	
280 285 290	
att tat cta gta att ctt ata tct gaa gct gtc atc agc act atc ttg	1087
Ile Tyr Leu Val Ile Leu Ile Ser Glu Ala Val Ile Ser Thr Ile Leu	
295 300 305	
aag tat aca tgg caa gct gaa gaa aaa tgg gat gaa cct tgg tat aac	1135
Lys Tyr Thr Trp Gln Ala Glu Glu Lys Trp Asp Glu Pro Trp Tyr Asn	
310 315 320	
caa aaa aca gaa cat caa aga aat agc agt aag att ctg aga ttt att	1183
Gln Lys Thr Glu His Gln Arg Asn Ser Ser Lys Ile Leu Arg Phe Ile	
325 330 335	
tca gac ttc ctt gct ttt ttg gtt ctc tac aat ttc atc att cca att	1231
Ser Asp Phe Leu Ala Phe Leu Val Leu Tyr Asn Phe Ile Ile Pro Ile	
340 345 350 355	
tca tta tat gtg aca gtc gaa atg cag aaa ttt ctt gga tca ttt ttt	1279
Ser Leu Tyr Val Thr Val Glu Met Gln Lys Phe Leu Gly Ser Phe Phe	
360 365 370	
att ggc tgg gat ctt gat ctg tat cat gaa gaa tca gat cag aaa gct	1327
Ile Gly Trp Asp Leu Asp Leu Tyr His Glu Glu Ser Asp Gln Lys Ala	
375 380 385	

caa gtc aat act tcc gat ctg aat gaa gag ctt gga cag gta gag tac	1375
Gln Val Asn Thr Ser Asp Leu Asn Glu Glu Leu Gly Gln Val Glu Tyr	
390 395 400	
gtg ttt aca gat aaa act ggt aca ctg aca gaa aat gag atg cag ttt	1423
Val Phe Thr Asp Lys Thr Gly Thr Leu Thr Glu Asn Glu Met Gln Phe	
405 410 415	
cgg gaa tgt tca att aat ggc atg aaa tac caa gaa att aat ggt aga	1471
Arg Glu Cys Ser Ile Asn Gly Met Lys Tyr Gln Glu Ile Asn Gly Arg	
420 425 430 435	
ctt gta ccc gaa gga cca aca cca gac tct tca gaa gga aac tta tct	1519
Leu Val Pro Glu Gly Pro Thr Pro Asp Ser Ser Glu Gly Asn Leu Ser	
440 445 450	
tat ctt agt agt tta tcc cat ctt aac aac tta tcc cat ctt aca acc	1567
Tyr Leu Ser Ser Leu Ser His Leu Asn Asn Leu Ser His Leu Thr Thr	
455 460 465	
agt tcc tct ttc aga acc agt cct gaa aat gaa act gaa cta att aaa	1615
Ser Ser Ser Phe Arg Thr Ser Pro Glu Asn Glu Thr Glu Leu Ile Lys	
470 475 480	
gaa cat gat ctc ttc ttt aaa gca gtc agt ctc tgt cac act gta cag	1663
Glu His Asp Leu Phe Phe Lys Ala Val Ser Leu Cys His Thr Val Gln	
485 490 495	
att agc aat gtt caa act gac tgc act ggt gat ggt ccc tgg caa tcc	1711
Ile Ser Asn Val Gln Thr Asp Cys Thr Gly Asp Gly Pro Trp Gln Ser	
500 505 510 515	
aac ctg gca cca tcc cag ttg gag tac tat gca tct tca cca gat gaa	1759
Asn Leu Ala Pro Ser Gln Leu Glu Tyr Tyr Ala Ser Ser Pro Asp Glu	
520 525 530	
aag gct cta gta gaa gct gct gca agg att ggt att gtg ttt att ggc	1807
Lys Ala Leu Val Glu Ala Ala Ala Arg Ile Gly Ile Val Phe Ile Gly	
535 540 545	
aat tct gaa gaa act atg gag gtt aaa act ctt gga aaa ctg gaa cgg	1855
Asn Ser Glu Glu Thr Met Glu Val Lys Thr Leu Gly Lys Leu Glu Arg	
550 555 560	
tac aaa ctg ctt cat att ctg gaa ttt gat tca gat cgt agg aga atg	1903
Tyr Lys Leu Leu His Ile Leu Glu Phe Asp Ser Asp Arg Arg Arg Met	
565 570 575	
agt gta att gtt cag gca cct tca ggt gag aag tta tta ttt gct aaa	1951
Ser Val Ile Val Gln Ala Pro Ser Gly Glu Lys Leu Leu Phe Ala Lys	
580 585 590 595	
gga gct gag tca tca att ctc cct aaa tgt ata ggt gga gaa ata gaa	1999
Gly Ala Glu Ser Ser Ile Leu Pro Lys Cys Ile Gly Gly Glu Ile Glu	
600 605 610	
aaa acc aga att cat gta gat gaa ttt gct ttg aaa ggg cta aga act	2047
Lys Thr Arg Ile His Val Asp Glu Phe Ala Leu Lys Gly Leu Arg Thr	
615 620 625	

ctg tgt ata gca tat aga aaa ttt aca tca aaa gag tat gag gaa ata	2095
Leu Cys Ile Ala Tyr Arg Lys Phe Thr Ser Lys Glu Tyr Glu Glu Ile	
630 635 640	
gat aaa cgc ata ttt gaa gcc agg act gcc ttg cag cag cgg gaa gag	2143
Asp Lys Arg Ile Phe Glu Ala Arg Thr Ala Leu Gln Gln Arg Glu Glu	
645 650 655	
aaa ttg gca gct gtt ttc cag ttc ata gag aaa gac ctg ata tta ctt	2191
Lys Leu Ala Ala Val Phe Gln Phe Ile Glu Lys Asp Leu Ile Leu Leu	
660 665 670 675	
gga gcc aca gca gta gaa gac aga cta caa gat aaa gtt cga gaa act	2239
Gly Ala Thr Ala Val Glu Asp Arg Leu Gln Asp Lys Val Arg Glu Thr	
680 685 690	
att gaa gca ttg aga atg gct ggt atc aaa gta tgg gta ctt act ggg	2287
Ile Glu Ala Leu Arg Met Ala Gly Ile Lys Val Trp Val Leu Thr Gly	
695 700 705	
gat aaa cat gaa aca gct gtt agt gtg agt tta tca tgt ggc cat ttt	2335
Asp Lys His Glu Thr Ala Val Ser Val Ser Leu Ser Cys Gly His Phe	
710 715 720	
cat aga acc atg aac atc ctt gaa ctt ata aac cag aaa tca gac agc	2383
His Arg Thr Met Asn Ile Leu Glu Leu Ile Asn Gln Lys Ser Asp Ser	
725 730 735	
gag tgt gct gaa caa ttg agg cag ctt gcc aga aga att aca gag gat	2431
Glu Cys Ala Glu Gln Leu Arg Gln Leu Ala Arg Arg Ile Thr Glu Asp	
740 745 750 755	
cat gtg att cag cat ggg ctg gta gtg gat ggg acc agc cta tct ctt	2479
His Val Ile Gln His Gly Leu Val Val Asp Gly Thr Ser Leu Ser Leu	
760 765 770	
gca ctc agg gag cat gaa aaa cta ttt atg gaa gtt tgc aga aat tgt	2527
Ala Leu Arg Glu His Glu Lys Leu Phe Met Glu Val Cys Arg Asn Cys	
775 780 785	
tca gct gta tta tgc tgt cgt atg gct cca ctg cag aaa gca aaa gta	2575
Ser Ala Val Leu Cys Cys Arg Met Ala Pro Leu Gln Lys Ala Lys Val	
790 795 800	
ata aga cta ata aaa ata tca cct gag aaa cct ata aca ttg gct gtt	2623
Ile Arg Leu Ile Lys Ile Ser Pro Glu Lys Pro Ile Thr Leu Ala Val	
805 810 815	
ggt gat ggt gct aat gac gta agc atg ata caa gaa gcc cat gtt ggc	2671
Gly Asp Gly Ala Asn Asp Val Ser Met Ile Gln Glu Ala His Val Gly	
820 825 830 835	
ata gga atc atg ggt aaa gaa gga aga cag gct gca aga aac agt gac	2719
Ile Gly Ile Met Gly Lys Glu Gly Arg Gln Ala Ala Arg Asn Ser Asp	
840 845 850	
tat gca ata gcc aga ttt aag ttc ctc tcc aaa ttg ctt ttt gtt cat	2767
Tyr Ala Ile Ala Arg Phe Lys Phe Leu Ser Lys Leu Leu Phe Val His	
855 860 865	

ggt cat ttt tat tat att aga ata gct acc ctt gta cag tat ttt ttt 2815
 Gly His Phe Tyr Tyr Ile Arg Ile Ala Thr Leu Val Gln Tyr Phe Phe
 870 875 880

tat aag aat gtg tgc ttt atc aca ccc cag ttt tta tat cag ttc tac 2863
 Tyr Lys Asn Val Cys Phe Ile Thr Pro Gln Phe Leu Tyr Gln Phe Tyr
 885 890 895

tgt ttg ttt tct cag caa aca ttg tat gac agc gtg tac ctg act tta 2911
 Cys Leu Phe Ser Gln Gln Thr Leu Tyr Asp Ser Val Tyr Leu Thr Leu
 900 905 910 915

tac aat att tgt ttt act tcc cta cct att ctg ata tat agt ctt ttg 2959
 Tyr Asn Ile Cys Phe Thr Ser Leu Pro Ile Leu Ile Tyr Ser Leu Leu
 920 925 930

gaa cag cat gta gac cct cat gtg tta caa aat aag ccc acc ctt tat 3007
 Glu Gln His Val Asp Pro His Val Leu Gln Asn Lys Pro Thr Leu Tyr
 935 940 945

cga gac att agt aaa aac cgc ctc tta agt att aaa aca ttt ctt tat 3055
 Arg Asp Ile Ser Lys Asn Arg Leu Ser Ile Lys Thr Phe Leu Tyr
 950 955 960

tgg acc atc ctg ggc ttc agt cat gcc ttt att ttc ttt ttt gga tcc 3103
 Trp Thr Ile Leu Gly Phe Ser His Ala Phe Ile Phe Phe Phe Gly Ser
 965 970 975

tat tta cta ata ggg aaa gat aca tct ctg ctt gga aat ggc cag atg 3151
 Tyr Leu Leu Ile Gly Lys Asp Thr Ser Leu Leu Gly Asn Gly Gln Met
 980 985 990 995

ttt gga aac tgg aca ttt ggc act ttg gtc ttc aca gtc atg gtt att 3199
 Phe Gly Asn Trp Thr Phe Gly Thr Leu Val Phe Thr Val Met Val Ile
 1000 1005 1010

aca gtc aca gta aag atg gct ctg gaa act cat ttt tgg act tgg atc 3247
 Thr Val Thr Val Lys Met Ala Leu Glu Thr His Phe Trp Thr Trp Ile
 1015 1020 1025

aac cat ctc gtt acc tgg gga tct att ata ttt tat ttt gta ttt tcc 3295
 Asn His Leu Val Thr Trp Gly Ser Ile Ile Phe Tyr Phe Val Phe Ser
 1030 1035 1040

ttg ttt tat gga ggg att ctc tgg coa ttt ttg ggc tcc cag aat atg 3343
 Leu Phe Tyr Gly Gly Ile Leu Trp Pro Phe Leu Gly Ser Gln Asn Met
 1045 1050 1055

tat ttt gtg ttt att cag ctc ctg tca agt ggt tct gct tgg ttt gcc 3391
 Tyr Phe Val Phe Ile Gln Leu Leu Ser Ser Gly Ser Ala Trp Phe Ala
 1060 1065 1070 1075

ata atc ctc atg gtt gtt aca tgt cta ttt ctt gat atc ata aag aag 3439
 Ile Ile Leu Met Val Val Thr Cys Leu Phe Leu Asp Ile Ile Lys Lys
 1080 1085 1090

gtc ttt gac cga cac ctc cac cct aca agt act gaa aag gca cag ctt 3487
 Val Phe Asp Arg His Leu His Pro Thr Ser Thr Glu Lys Ala Gln Leu
 1095 1100 1105

act gaa aca aat gca ggt atc aag tgc ttg gac tcc atg tgc tgt ttc 3535
 Thr Glu Thr Asn Ala Gly Ile Lys Cys Leu Asp Ser Met Cys Cys Phe
 1110 1115 1120

ccg gaa gga gaa gca gcg tgt gca tct gtt gga aga atg ctg gaa cga 3583
 Pro Glu Gly Glu Ala Ala Cys Ala Ser Val Gly Arg Met Leu Glu Arg
 1125 1130 1135

gtt ata gga aga tgt agt cca acc cac atc agc agt tca tgg agt gca 3631
 Val Ile Gly Arg Cys Ser Pro Thr His Ile Ser Ser Ser Trp Ser Ala
 1140 1145 1150 1155

tgc gat cct ttc tat acc aac gac agg agc atc ttg act ctc tcc aca 3679
 Ser Asp Pro Phe Tyr Thr Asn Asp Arg Ser Ile Leu Thr Leu Ser Thr
 1160 1165 1170

atg gac tca tct act tgt taa aggggcagta gtactttgtg ggagccagtt 3730
 Met Asp Ser Ser Thr Cys *
 1175

cacctccttt cctaaaaattc agtgtgatca ccctgttaat ggccacacta gctctgaaat 3790
 taatttccaa aatctttgtg gtagttcata ccactcaga gttataatgg caaacaaca 3850
 gaaagcatta gtacaagccc ctcccacac ccttaatttg aatctgaaca tgttaaaatt 3910
 tgagaataaa gagacatttt tcatctcttt gtctggttg tcccttggtc ttatgggact 3970
 cctaattggca tttagtctg ttgctgagc cattatatt taatataaat gtagaaaaaa 4030
 gagagaaatc ttagtaaaag gtatttttta gtattagctt gattattgac tcttctattt 4090
 aaatctgctt ctgtaaatga tgctgaaagt ttgccttgag aactctattt ttttattaga 4150
 gttatatatta aagcttttca tgggaaaagt taatgtgaat actgaggaat tttggctcct 4210
 cagtgcctg tgttgtaaat tcattaatgc attctgagtt cacagagcaa attaggagaa 4270
 tcatttccaa ccattattta ctgcagtatg gggagtaaat ttataccaat tcctctaact 4330
 gtactgtaac acagcctgta aagttagcca tataaatgca aggtatatac atatatacaa 4390
 atcaggaatc aggtccgttc accgaacttc aaattgatgt ttactaatat tttgtgaca 4450
 gagtataaag accctatagt gggtaaatga gatactatta gcatattatt aatttaattg 4510
 ctttatcatt ggatcttttg catgctttaa tctggttaac atatttaaat ttgctttttt 4570
 tctctttacc tgaaggctct gtgtatagta ttcatgaca tctgtgtaca gtttaactat 4630
 caataaaaag tttggacagt atttaaatat tgcaaatatg ttttaattata caaatcagaa 4690
 tagtatgggt aattaaatga atacaaaag aagagcctct ttctgcagcc gacttagaca 4750
 tgctcttccc tttctataag ctagatttta gaataaaggg tttcagttaa taatcttatt 4810
 ttcaggttat gtcatttaac ttatagcaaa ctaccacaat acagttaggt ctgccaggtg 4870
 cccagtacaa ggcatttttc aggtgtggct gtggaatgta aaaatgctca acttgatca 4930
 ggtaatgta gcaataaatt aaatgctaag aatgattaat cgggtacatg ttactgtaat 4990
 taactcattg cacttcaaaa cctaacttcc atcctgaatt tatcaagtag tttagtttg 5050
 tcatttgttt ttgttttatt gaaaagtaat gttgtcttaa gatttagaag tgattattag 5110
 cttgagaact attaccagc tctaagcaaa taatgattgt atacatatta agataatggt 5170
 taaatgcggt tttaaccaag tttcccttga aaatgtaatt cctttatgga gatttattgt 5230
 gcagccctaa gcttccctcc catttcatga atataaggct tctagaattg gactggcagg 5290
 ggaaagaatg gtagagacag aaattaagac tttatccttg tttgcttgta aactattatt 5350
 ttcttgctaa tgtaacattt gtctgttcca gtgatgtaag gatattaagt tattaagcta 5410
 aatattaatt ttcaaaaata gtccttcttt aacttagata ttcatagct ggatttagga 5470
 agatctgtta ttctggaagt actaaaaaga ataatacaac gtacaatgtc tgcattcact 5530
 aattcatgtt ccagaagagg aaataatgaa gatatactca gtagagtact aggtgggagg 5590
 atatggaaat ttgctcataa aatctcttat aaaacgtgca tataacaaa tgacaccag 5650
 taggcctgca ttacatttac atgaccgtgt ttatttgcca tcaataaaac tgagtactga 5710
 caccagacaa agactccaaa gtcataaaat agcctatgac caactgcagc aagacaggag 5770
 gtcagctgc ctataatggt gcttaagtgt tgattgatgt aattttctgt actcaccatt 5830
 tgaagttagt taaggagaac tttatttttt taaaaaagt aaatggcaac cactagtgtg 5890
 ctcatcctga actgttactc caaatccact cgttttttaa agcaaaatta tcttgtgatt 5950
 ttaagaaaag agttttctat ttatttaaga aagtaacaat gcagtctgca agctttcagt 6010
 agttttctag tgctatattc atcctgtaaa actcttacta cgtaaccagt aatcacaagg 6070
 aaagtgtccc ctttgcatat ttcttttaaa ttctttcttt ggaaagtatg atgttgataa 6130
 ttaacttacc cttatctgcc aaaaccagag caaatgcta aatacgttat tgctaactcag 6190

```

tgggtctcaaaa tcgatttgcc tccctttgcc tcgtctgagg gctgtaagcc tgaagatagt 6250
ggcaagcacc aagtcagttt ccaaaattgc ccctcagctg ctttaagtga ctcagcagccc 6310
tgcctcagct tcagcaggcg taggtccacc ctgggcggag caaagtatgg gccagggaga 6370
actacagcta cgaagaactg ctgtcgagtt gagaanaaggg gagaatttat ggtctgatt 6430
ttctaactgt cctctttctt gggctctaaag ctcataatac acaaaggctt ccagacctga 6490
gccacacca ggcctatcc tgaacaggag actaaacaga ggcaaataca ccctaggaaa 6550
tacttgcat ctgccctacg gttagtacca ggactgaggt catttctact ggaaaagatt 6610
gtgagattga acttatctga tcgcttgaga ctccataatag gcaggagtca aggccactag 6670
aaaattgaca gtttaagagcc aaaagttttt aaaatatgct actctgaaaa atctcgtgaa 6730
ggctgtagga aaaggagaa tcttccatgt tgggtgtttt cctgtaaaaga tcagtttggg 6790
gtatgatata agcaggtatt aataaaaaa acacacaaa gagttacgta aaacatgttt 6850
tattaatttt ggtccccacg tacagacatt ttatttctat ttgaaatga gttatctatt 6910
ttcataaaaag taaaacacta ttaaagtgtc gttttatgtg aaataacttg aatgttggtc 6970
ctataaaaaa tagatcataa ctcatgatat gtttgtaatc atggtaattt agatttttat 7030
gaggaatgag tatctggaaa tattgtagca atacttggtt taaaattttg gacctgagac 7090
actgtggctg tctaattgta tcccttaaaa attctctgca ttgtcagtaa atgtagtata 7150
ttattgtaca gctactcata attttttaaa gtttatgaag ttatatttat caaataaaaa 7210
ctttcctata t 7221

```

<210> 20

<211> 1177

<212> PRT

<213> Homo sapiens

<400> 20

```

Met Trp Arg Trp Ile Arg Gln Gln Leu Gly Phe Asp Pro Pro His Gln
1      5      10      15
Ser Asp Thr Arg Thr Ile Tyr Val Ala His Arg Phe Pro Gln Asn Gly
20      25      30
Leu Tyr Thr Pro Gln Lys Phe Ile Asp Asn Arg Ile Ile Ser Ser Lys
35      40      45
Tyr Thr Val Trp Asn Phe Val Pro Lys Asn Leu Phe Glu Gln Phe Arg
50      55      60
Arg Val Ala Asn Phe Tyr Phe Leu Ile Ile Phe Leu Val Gln Leu Met
65      70      75      80
Ile Asp Thr Pro Thr Ser Pro Val Thr Ser Gly Leu Pro Leu Phe Phe
85      90      95
Val Ile Thr Val Thr Ala Ile Lys Gln Gly Tyr Glu Asp Trp Leu Arg
100     105     110
His Asn Ser Asp Asn Glu Val Asn Gly Ala Pro Val Tyr Val Val Arg
115     120     125
Ser Gly Gly Leu Val Lys Thr Arg Ser Lys Asn Ile Arg Val Gly Asp
130     135     140
Ile Val Arg Ile Ala Lys Asp Glu Ile Phe Pro Ala Asp Leu Val Leu
145     150     155     160
Leu Ser Ser Asp Arg Leu Asp Gly Ser Cys His Val Thr Thr Ala Ser
165     170     175
Leu Asp Gly Glu Thr Asn Leu Lys Thr His Val Ala Val Pro Glu Thr
180     185     190
Ala Leu Leu Gln Thr Val Ala Asn Leu Asp Thr Leu Val Ala Val Ile
195     200     205
Glu Cys Gln Gln Pro Glu Ala Asp Leu Tyr Arg Phe Met Gly Arg Met
210     215     220
Ile Ile Thr Gln Gln Met Glu Glu Ile Val Arg Pro Leu Gly Pro Glu
225     230     235     240
Ser Leu Leu Leu Arg Gly Ala Arg Leu Lys Asn Thr Lys Glu Ile Phe
245     250     255
Gly Val Ala Val Tyr Thr Gly Met Glu Thr Lys Met Ala Leu Asn Tyr
260     265     270

```

Lys	Ser	Lys	Ser	Gln	Lys	Arg	Ser	Ala	Val	Glu	Lys	Ser	Met	Asn	Thr	275	280	285
Phe	Leu	Ile	Ile	Tyr	Leu	Val	Ile	Leu	Ile	Ser	Glu	Ala	Val	Ile	Ser	290	295	300
Thr	Ile	Leu	Lys	Tyr	Thr	Trp	Gln	Ala	Glu	Glu	Lys	Trp	Asp	Glu	Pro	305	310	315
Trp	Tyr	Asn	Gln	Lys	Thr	Glu	His	Gln	Arg	Asn	Ser	Ser	Lys	Ile	Leu	325	330	335
Arg	Phe	Ile	Ser	Asp	Phe	Leu	Ala	Phe	Leu	Val	Leu	Tyr	Asn	Phe	Ile	340	345	350
Ile	Pro	Ile	Ser	Leu	Tyr	Val	Thr	Val	Glu	Met	Gln	Lys	Phe	Leu	Gly	355	360	365
Ser	Phe	Phe	Ile	Gly	Trp	Asp	Leu	Asp	Leu	Tyr	His	Glu	Glu	Ser	Asp	370	375	380
Gln	Lys	Ala	Gln	Val	Asn	Thr	Ser	Asp	Leu	Asn	Glu	Glu	Leu	Gly	Gln	385	390	395
Val	Glu	Tyr	Val	Phe	Thr	Asp	Lys	Thr	Gly	Thr	Leu	Thr	Glu	Asn	Glu	405	410	415
Met	Gln	Phe	Arg	Glu	Cys	Ser	Ile	Asn	Gly	Met	Lys	Tyr	Gln	Glu	Ile	420	425	430
Asn	Gly	Arg	Leu	Val	Pro	Glu	Gly	Pro	Thr	Pro	Asp	Ser	Ser	Glu	Gly	435	440	445
Asn	Leu	Ser	Tyr	Leu	Ser	Ser	Leu	Ser	His	Leu	Asn	Asn	Leu	Ser	His	450	455	460
Leu	Thr	Thr	Ser	Ser	Ser	Phe	Arg	Thr	Ser	Pro	Glu	Asn	Glu	Thr	Glu	465	470	475
Leu	Ile	Lys	Glu	His	Asp	Leu	Phe	Phe	Lys	Ala	Val	Ser	Leu	Cys	His	485	490	495
Thr	Val	Gln	Ile	Ser	Asn	Val	Gln	Thr	Asp	Cys	Thr	Gly	Asp	Gly	Pro	500	505	510
Trp	Gln	Ser	Asn	Leu	Ala	Pro	Ser	Gln	Leu	Glu	Tyr	Tyr	Ala	Ser	Ser	515	520	525
Pro	Asp	Glu	Lys	Ala	Leu	Val	Glu	Ala	Ala	Ala	Arg	Ile	Gly	Ile	Val	530	535	540
Phe	Ile	Gly	Asn	Ser	Glu	Glu	Thr	Met	Glu	Val	Lys	Thr	Leu	Gly	Lys	545	550	555
Leu	Glu	Arg	Tyr	Lys	Leu	Leu	His	Ile	Leu	Glu	Phe	Asp	Ser	Asp	Arg	565	570	575
Arg	Arg	Met	Ser	Val	Ile	Val	Gln	Ala	Pro	Ser	Gly	Glu	Lys	Leu	Leu	580	585	590
Phe	Ala	Lys	Gly	Ala	Glu	Ser	Ser	Ile	Leu	Pro	Lys	Cys	Ile	Gly	Gly	595	600	605
Glu	Ile	Glu	Lys	Thr	Arg	Ile	His	Val	Asp	Glu	Phe	Ala	Leu	Lys	Gly	610	615	620
Leu	Arg	Thr	Leu	Cys	Ile	Ala	Tyr	Arg	Lys	Phe	Thr	Ser	Lys	Glu	Tyr	625	630	635
Glu	Glu	Ile	Asp	Lys	Arg	Ile	Phe	Glu	Ala	Arg	Thr	Ala	Leu	Gln	Gln	645	650	655
Arg	Glu	Glu	Lys	Leu	Ala	Ala	Val	Phe	Gln	Phe	Ile	Glu	Lys	Asp	Leu	660	665	670
Ile	Leu	Leu	Gly	Ala	Thr	Ala	Val	Glu	Asp	Arg	Leu	Gln	Asp	Lys	Val	675	680	685
Arg	Glu	Thr	Ile	Glu	Ala	Leu	Arg	Met	Ala	Gly	Ile	Lys	Val	Trp	Val	690	695	700
Leu	Thr	Gly	Asp	Lys	His	Glu	Thr	Ala	Val	Ser	Val	Ser	Leu	Ser	Cys	705	710	715
Gly	His	Phe	His	Arg	Thr	Met	Asn	Ile	Leu	Glu	Leu	Ile	Asn	Gln	Lys	725	730	735
Ser	Asp	Ser	Glu	Cys	Ala	Glu	Gln	Leu	Arg	Gln	Leu	Ala	Arg	Arg	Ile	740	745	750

Thr	Glu	Asp	His	Val	Ile	Gln	His	Gly	Leu	Val	Val	Asp	Gly	Thr	Ser
		755					760					765			
Leu	Ser	Leu	Ala	Leu	Arg	Glu	His	Glu	Lys	Leu	Phe	Met	Glu	Val	Cys
		770				775					780				
Arg	Asn	Cys	Ser	Ala	Val	Leu	Cys	Cys	Arg	Met	Ala	Pro	Leu	Gln	Lys
785					790					795					800
Ala	Lys	Val	Ile	Arg	Leu	Ile	Lys	Ile	Ser	Pro	Glu	Lys	Pro	Ile	Thr
				805					810					815	
Leu	Ala	Val	Gly	Asp	Gly	Ala	Asn	Asp	Val	Ser	Met	Ile	Gln	Glu	Ala
			820					825					830		
His	Val	Gly	Ile	Gly	Ile	Met	Gly	Lys	Glu	Gly	Arg	Gln	Ala	Ala	Arg
		835					840					845			
Asn	Ser	Asp	Tyr	Ala	Ile	Ala	Arg	Phe	Lys	Phe	Leu	Ser	Lys	Leu	Leu
		850				855					860				
Phe	Val	His	Gly	His	Phe	Tyr	Tyr	Ile	Arg	Ile	Ala	Thr	Leu	Val	Gln
865					870				875						880
Tyr	Phe	Phe	Tyr	Lys	Asn	Val	Cys	Phe	Ile	Thr	Pro	Gln	Phe	Leu	Tyr
				885					890					895	
Gln	Phe	Tyr	Cys	Leu	Phe	Ser	Gln	Gln	Thr	Leu	Tyr	Asp	Ser	Val	Tyr
			900					905					910		
Leu	Thr	Leu	Tyr	Asn	Ile	Cys	Phe	Thr	Ser	Leu	Pro	Ile	Leu	Ile	Tyr
			915					920				925			
Ser	Leu	Leu	Glu	Gln	His	Val	Asp	Pro	His	Val	Leu	Gln	Asn	Lys	Pro
			930			935					940				
Thr	Leu	Tyr	Arg	Asp	Ile	Ser	Lys	Asn	Arg	Leu	Leu	Ser	Ile	Lys	Thr
945					950					955					960
Phe	Leu	Tyr	Trp	Thr	Ile	Leu	Gly	Phe	Ser	His	Ala	Phe	Ile	Phe	Phe
				965					970					975	
Phe	Gly	Ser	Tyr	Leu	Leu	Ile	Gly	Lys	Asp	Thr	Ser	Leu	Leu	Gly	Asn
			980					985					990		
Gly	Gln	Met	Phe	Gly	Asn	Trp	Thr	Phe	Gly	Thr	Leu	Val	Phe	Thr	Val
			995				1000					1005			
Met	Val	Ile	Thr	Val	Thr	Val	Lys	Met	Ala	Leu	Glu	Thr	His	Phe	Trp
			1010			1015					1020				
Thr	Trp	Ile	Asn	His	Leu	Val	Thr	Trp	Gly	Ser	Ile	Ile	Phe	Tyr	Phe
1025					1030					1035					1040
Val	Phe	Ser	Leu	Phe	Tyr	Gly	Gly	Ile	Leu	Trp	Pro	Phe	Leu	Gly	Ser
				1045					1050					1055	
Gln	Asn	Met	Tyr	Phe	Val	Phe	Ile	Gln	Leu	Leu	Ser	Ser	Gly	Ser	Ala
			1060					1065					1070		
Trp	Phe	Ala	Ile	Ile	Leu	Met	Val	Val	Thr	Cys	Leu	Phe	Leu	Asp	Ile
			1075				1080						1085		
Ile	Lys	Lys	Val	Phe	Asp	Arg	His	Leu	His	Pro	Thr	Ser	Thr	Glu	Lys
			1090			1095					1100				
Ala	Gln	Leu	Thr	Glu	Thr	Asn	Ala	Gly	Ile	Lys	Cys	Leu	Asp	Ser	Met
1105					1110					1115					1120
Cys	Cys	Phe	Pro	Glu	Gly	Glu	Ala	Ala	Cys	Ala	Ser	Val	Gly	Arg	Met
				1125					1130					1135	
Leu	Glu	Arg	Val	Ile	Gly	Arg	Cys	Ser	Pro	Thr	His	Ile	Ser	Ser	Ser
			1140					1145					1150		
Trp	Ser	Ala	Ser	Asp	Pro	Phe	Tyr	Thr	Asn	Asp	Arg	Ser	Ile	Leu	Thr
			1155				1160						1165		
Leu	Ser	Thr	Met	Asp	Ser	Ser	Thr	Cys							
			1170				1175								

<210> 21

<211> 3534

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)... (3534)

<400> 21

atg tgg cgg tgg atc cgg cag cag ctg ggt ttt gac cca cca cat cag	48
Met Trp Arg Trp Ile Arg Gln Gln Leu Gly Phe Asp Pro Pro His Gln	
1 5 10 15	
agt gac aca aga acc atc tac gta gcc cac agg ttt cct cag aat ggc	96
Ser Asp Thr Arg Thr Ile Tyr Val Ala His Arg Phe Pro Gln Asn Gly	
20 25 30	
ctt tac aca cct cag aaa ttt ata gat aac agg atc att tca tct aag	144
Leu Tyr Thr Pro Gln Lys Phe Ile Asp Asn Arg Ile Ile Ser Ser Lys	
35 40 45	
tac act gtg tgg aat ttt gtt cca aaa aat tta ttt gaa cag ttc aga	192
Tyr Thr Val Trp Asn Phe Val Pro Lys Asn Leu Phe Glu Gln Phe Arg	
50 55 60	
aga gtg gca aac ttt tat ttt ctt att ata ttt ttg gtt cag ctt atg	240
Arg Val Ala Asn Phe Tyr Phe Leu Ile Ile Phe Leu Val Gln Leu Met	
65 70 75 80	
att gat aca cct acc agt cca gtt acc agt gga ctt cca tta ttc ttt	288
Ile Asp Thr Pro Thr Ser Pro Val Thr Ser Gly Leu Pro Leu Phe Phe	
85 90 95	
gtg ata aca gta act gcc ata aag cag gga tat gaa gat tgg tta cgg	336
Val Ile Thr Val Thr Ala Ile Lys Gln Gly Tyr Glu Asp Trp Leu Arg	
100 105 110	
cat aac tca gat aat gaa gta aat gga gct cct gtt tat gtt gtt cga	384
His Asn Ser Asp Asn Glu Val Asn Gly Ala Pro Val Tyr Val Val Arg	
115 120 125	
agt ggt ggc ctt gta aaa act aga tca aaa aac att cgg gtg ggt gat	432
Ser Gly Gly Leu Val Lys Thr Arg Ser Lys Asn Ile Arg Val Gly Asp	
130 135 140	
att gtt cga ata gcc aaa gat gaa att ttt cct gca gac ttg gtg ctt	480
Ile Val Arg Ile Ala Lys Asp Glu Ile Phe Pro Ala Asp Leu Val Leu	
145 150 155 160	
ctg tcc tca gat cga ctg gat ggt tcc tgt cac gtt aca act gct agt	528
Leu Ser Ser Asp Arg Leu Asp Gly Ser Cys His Val Thr Thr Ala Ser	
165 170 175	
ttg gac gga gaa act aac ctg aag aca cat gtg gca gtt cca gaa aca	576
Leu Asp Gly Glu Thr Asn Leu Lys Thr His Val Ala Val Pro Glu Thr	
180 185 190	
gca tta tta caa aca gtt gcc aat ttg gac act cta gta gct gta ata	624
Ala Leu Leu Gln Thr Val Ala Asn Leu Asp Thr Leu Val Ala Val Ile	
195 200 205	
gaa tgc cag caa cca gaa gca gac tta tac aga ttc atg gga cga atg	672
Glu Cys Gln Gln Pro Glu Ala Asp Leu Tyr Arg Phe Met Gly Arg Met	
210 215 220	

atc	ata	acc	caa	caa	atg	gaa	gaa	att	gta	aga	cct	ctg	ggg	ccg	gag	720
Ile	Ile	Thr	Gln	Gln	Met	Glu	Glu	Ile	Val	Arg	Pro	Leu	Gly	Pro	Glu	
225					230				235						240	
agt	ctc	ctg	ctt	cgt	gga	gcc	aga	tta	aaa	aac	aca	aaa	gaa	att	ttt	768
Ser	Leu	Leu	Leu	Arg	Gly	Ala	Arg	Leu	Lys	Asn	Thr	Lys	Glu	Ile	Phe	
				245				250						255		
ggt	gtt	gcg	gta	tac	act	gga	atg	gaa	act	aag	atg	gca	tta	aat	tac	816
Gly	Val	Ala	Val	Tyr	Thr	Gly	Met	Glu	Thr	Lys	Met	Ala	Leu	Asn	Tyr	
			260					265					270			
aag	agc	aaa	tca	cag	aaa	cga	tct	gca	gta	gaa	aag	tca	atg	aat	aca	864
Lys	Ser	Lys	Ser	Gln	Lys	Arg	Ser	Ala	Val	Glu	Lys	Ser	Met	Asn	Thr	
		275					280					285				
ttt	ttg	ata	att	tat	cta	gta	att	ctt	ata	tct	gaa	gct	gtc	atc	agc	912
Phe	Leu	Ile	Ile	Tyr	Leu	Val	Ile	Leu	Ile	Ser	Glu	Ala	Val	Ile	Ser	
	290					295					300					
act	atc	ttg	aag	tat	aca	tgg	caa	gct	gaa	gaa	aaa	tgg	gat	gaa	cct	960
Thr	Ile	Leu	Lys	Tyr	Thr	Trp	Gln	Ala	Glu	Glu	Lys	Trp	Asp	Glu	Pro	
305					310				315						320	
tgg	tat	aac	caa	aaa	aca	gaa	cat	caa	aga	aat	agc	agt	aag	att	ctg	1008
Trp	Tyr	Asn	Gln	Lys	Thr	Glu	His	Gln	Arg	Asn	Ser	Ser	Lys	Ile	Leu	
			325					330						335		
aga	ttt	att	tca	gac	ttc	ctt	gct	ttt	ttg	gtt	ctc	tac	aat	ttc	atc	1056
Arg	Phe	Ile	Ser	Asp	Phe	Leu	Ala	Phe	Leu	Val	Leu	Tyr	Asn	Phe	Ile	
			340					345					350			
att	cca	att	tca	tta	tat	gtg	aca	gtc	gaa	atg	cag	aaa	ttt	ctt	gga	1104
Ile	Pro	Ile	Ser	Leu	Tyr	Val	Thr	Val	Glu	Met	Gln	Lys	Phe	Leu	Gly	
	355					360					365					
tca	ttt	ttt	att	ggc	tgg	gat	ctt	gat	ctg	tat	cat	gaa	gaa	tca	gat	1152
Ser	Phe	Phe	Ile	Gly	Trp	Asp	Leu	Asp	Leu	Tyr	His	Glu	Glu	Ser	Asp	
	370					375					380					
cag	aaa	gct	caa	gtc	aat	act	tcc	gat	ctg	aat	gaa	gag	ctt	gga	cag	1200
Gln	Lys	Ala	Gln	Val	Asn	Thr	Ser	Asp	Leu	Asn	Glu	Glu	Leu	Gly	Gln	
385					390					395					400	
gta	gag	tac	gtg	ttt	aca	gat	aaa	aot	ggt	aca	ctg	aca	gaa	aat	gag	1248
Val	Glu	Tyr	Val	Phe	Thr	Asp	Lys	Thr	Gly	Thr	Leu	Thr	Glu	Asn	Glu	
			405					410					415			
atg	cag	ttt	cgg	gaa	tgt	tca	att	aat	ggc	atg	aaa	tac	caa	gaa	att	1296
Met	Gln	Phe	Arg	Glu	Cys	Ser	Ile	Asn	Gly	Met	Lys	Tyr	Gln	Glu	Ile	
			420					425					430			
aat	ggt	aga	ctt	gta	ccc	gaa	gga	cca	aca	cca	gac	tct	tca	gaa	gga	1344
Asn	Gly	Arg	Leu	Val	Pro	Glu	Gly	Pro	Thr	Pro	Asp	Ser	Ser	Glu	Gly	
			435				440					445				
aac	tta	tct	tat	ctt	agt	agt	tta	tcc	cat	ctt	aac	aac	tta	tcc	cat	1392
Asn	Leu	Ser	Tyr	Leu	Ser	Ser	Leu	Ser	His	Leu	Asn	Asn	Leu	Ser	His	
	450					455					460					

ctt	aca	acc	agt	tcc	tct	ttc	aga	acc	agt	cct	gaa	aat	gaa	act	gaa	1440
Leu	Thr	Thr	Ser	Ser	Ser	Phe	Arg	Thr	Ser	Pro	Glu	Asn	Glu	Thr	Glu	
465					470					475					480	
cta	att	aaa	gaa	cat	gat	ctc	ttc	ttt	aaa	gca	gtc	agt	ctc	tgt	cac	1488
Leu	Ile	Lys	Glu	His	Asp	Leu	Phe	Phe	Lys	Ala	Val	Ser	Leu	Cys	His	
				485					490					495		
act	gta	cag	att	agc	aat	gtt	caa	act	gac	tgc	act	ggg	gat	ggg	ccc	1536
Thr	Val	Gln	Ile	Ser	Asn	Val	Gln	Thr	Asp	Cys	Thr	Gly	Asp	Gly	Pro	
			500					505					510			
tgg	caa	tcc	aac	ctg	gca	cca	tgc	cag	ttg	gag	tac	tat	gca	tct	tca	1584
Trp	Gln	Ser	Asn	Leu	Ala	Pro	Ser	Gln	Leu	Glu	Tyr	Tyr	Ala	Ser	Ser	
		515					520					525				
cca	gat	gaa	aag	gct	cta	gta	gaa	gct	gct	gca	agg	att	ggg	att	gtg	1632
Pro	Asp	Glu	Lys	Ala	Leu	Val	Glu	Ala	Ala	Ala	Arg	Ile	Gly	Ile	Val	
	530					535					540					
ttt	att	ggc	aat	tct	gaa	gaa	act	atg	gag	gtt	aaa	act	ctt	gga	aaa	1680
Phe	Ile	Gly	Asn	Ser	Glu	Glu	Thr	Met	Glu	Val	Lys	Thr	Leu	Gly	Lys	
545					550					555					560	
ctg	gaa	cgg	tac	aaa	ctg	ctt	cat	att	ctg	gaa	ttt	gat	tca	gat	cgt	1728
Leu	Glu	Arg	Tyr	Lys	Leu	Leu	His	Ile	Leu	Glu	Phe	Asp	Ser	Asp	Arg	
				565					570					575		
agg	aga	atg	agt	gta	att	gtt	cag	gca	cct	tca	ggg	gag	aag	tta	tta	1776
Arg	Arg	Met	Ser	Val	Ile	Val	Gln	Ala	Pro	Ser	Gly	Glu	Lys	Leu	Leu	
			580					585					590			
ttt	gct	aaa	gga	gct	gag	tca	tca	att	ctc	cct	aaa	tgt	ata	ggg	gga	1824
Phe	Ala	Lys	Gly	Ala	Glu	Ser	Ser	Ile	Leu	Pro	Lys	Cys	Ile	Gly	Gly	
		595					600					605				
gaa	ata	gaa	aaa	acc	aga	att	cat	gta	gat	gaa	ttt	gct	ttg	aaa	ggg	1872
Glu	Ile	Glu	Lys	Thr	Arg	Ile	His	Val	Asp	Glu	Phe	Ala	Leu	Lys	Gly	
	610					615					620					
cta	aga	act	ctg	tgt	ata	gca	tat	aga	aaa	ttt	aca	tca	aaa	gag	tat	1920
Leu	Arg	Thr	Leu	Cys	Ile	Ala	Tyr	Arg	Lys	Phe	Thr	Ser	Lys	Glu	Tyr	
625					630					635					640	
gag	gaa	ata	gat	aaa	cgc	ata	ttt	gaa	gcc	agg	act	gcc	ttg	cag	cag	1968
Glu	Glu	Ile	Asp	Lys	Arg	Ile	Phe	Glu	Ala	Arg	Thr	Ala	Leu	Gln	Gln	
				645					650					655		
cgg	gaa	gag	aaa	ttg	gca	gct	gtt	ttc	cag	ttc	ata	gag	aaa	gac	ctg	2016
Arg	Glu	Glu	Lys	Leu	Ala	Ala	Val	Phe	Gln	Phe	Ile	Glu	Lys	Asp	Leu	
			660					665					670			
ata	tta	ctt	gga	gcc	aca	gca	gta	gaa	gac	aga	cta	caa	gat	aaa	gtt	2064
Ile	Leu	Leu	Gly	Ala	Thr	Ala	Val	Glu	Asp	Arg	Leu	Gln	Asp	Lys	Val	
		675					680					685				
cga	gaa	act	att	gaa	gca	ttg	aga	atg	gct	ggg	atc	aaa	gta	tgg	gta	2112
Arg	Glu	Thr	Ile	Glu	Ala	Leu	Arg	Met	Ala	Gly	Ile	Lys	Val	Trp	Val	
	690					695					700					

ctt act ggg gat aaa cat gaa aca gct gtt agt gtg agt tta tca tgt	2160
Leu Thr Gly Asp Lys His Glu Thr Ala Val Ser Val Ser Leu Ser Cys	
705 710 715 720	
ggc cat ttt cat aga acc atg aac atc ctt gaa ctt ata aac cag aaa	2208
Gly His Phe His Arg Thr Met Asn Ile Leu Glu Leu Ile Asn Gln Lys	
725 730 735	
tca gac agc gag tgt gct gaa caa ttg agg cag ctt gcc aga aga att	2256
Ser Asp Ser Glu Cys Ala Glu Gln Leu Arg Gln Leu Ala Arg Arg Ile	
740 745 750	
aca gag gat cat gtg att cag cat ggg ctg gta gtg gat ggg acc agc	2304
Thr Glu Asp His Val Ile Gln His Gly Leu Val Val Asp Gly Thr Ser	
755 760 765	
cta tct ctt gca ctc agg gag cat gaa aaa cta ttt atg gaa gtt tgc	2352
Leu Ser Leu Ala Leu Arg Glu His Glu Lys Leu Phe Met Glu Val Cys	
770 775 780	
aga aat tgt tca gct gta tta tgc tgt cgt atg gct cca ctg cag aaa	2400
Arg Asn Cys Ser Ala Val Leu Cys Cys Arg Met Ala Pro Leu Gln Lys	
785 790 795 800	
gca aaa gta ata aga cta ata aaa ata tca cct gag aaa cct ata acg	2448
Ala Lys Val Ile Arg Leu Ile Lys Ile Ser Pro Glu Lys Pro Ile Thr	
805 810 815	
ttg gct gtt ggt gat ggt gct aat gac gta agc atg ata caa gaa gcc	2496
Leu Ala Val Gly Asp Gly Ala Asn Asp Val Ser Met Ile Gln Glu Ala	
820 825 830	
cat gtt ggc ata gga atc atg ggt aaa gaa gga aga cag gct gca aga	2544
His Val Gly Ile Gly Ile Met Gly Lys Glu Gly Arg Gln Ala Ala Arg	
835 840 845	
aac agt gac tat gca ata gcc aga ttt aag ttc ctc tcc aaa ttg ctt	2592
Asn Ser Asp Tyr Ala Ile Ala Arg Phe Lys Phe Leu Ser Lys Leu Leu	
850 855 860	
ttt gtt cat ggt cat ttt tat tat att aga ata gct acc ctt gta cag	2640
Phe Val His Gly His Phe Tyr Tyr Ile Arg Ile Ala Thr Leu Val Gln	
865 870 875 880	
tat ttt ttt tat aag aat gtg tgc ttt atc aca ccc cag ttt tta tat	2688
Tyr Phe Phe Tyr Lys Asn Val Cys Phe Ile Thr Pro Gln Phe Leu Tyr	
885 890 895	
cag ttc tac tgt ttg ttt tct cag caa aca ttg tat gac agc gtg tac	2736
Gln Phe Tyr Cys Leu Phe Ser Gln Gln Thr Leu Tyr Asp Ser Val Tyr	
900 905 910	
ctg act tta tac aat att tgt ttt act tcc cta cct att ctg ata tat	2784
Leu Thr Leu Tyr Asn Ile Cys Phe Thr Ser Leu Pro Ile Leu Ile Tyr	
915 920 925	
agt ctt ttg gaa cag cat gta gac cct cat gtg tta caa aat aag ccc	2832
Ser Leu Leu Glu Gln His Val Asp Pro His Val Leu Gln Asn Lys Pro	
930 935 940	

acc ctt tat cga gac att agt aaa aac cgc ctg tta agt att aaa aca	2880
Thr Leu Tyr Arg Asp Ile Ser Lys Asn Arg Leu Leu Ser Ile Lys Thr	
945 950 955 960	
ttt ctt tat tgg acc atc ctg ggc ttc agt cat gcc ttt att ttc ttt	2928
Phe Leu Tyr Trp Thr Ile Leu Gly Phe Ser His Ala Phe Ile Phe Phe	
965 970 975	
ttt gga tcc tat tta cta ata ggg aaa gat aca tct ctg ctt gga aat	2976
Phe Gly Ser Tyr Leu Leu Ile Gly Lys Asp Thr Ser Leu Leu Gly Asn	
980 985 990	
ggc cag atg ttt gga aac tgg aca ttt ggc act ttg gtc ttc aca gtc	3024
Gly Gln Met Phe Gly Asn Trp Thr Phe Gly Thr Leu Val Phe Thr Val	
995 1000 1005	
atg gtt att aca gtc aca gta aag atg gct ctg gaa act cat ttt tgg	3072
Met Val Ile Thr Val Thr Val Lys Met Ala Leu Glu Thr His Phe Trp	
1010 1015 1020	
act tgg atc aac cat ctg gtt acc tgg gga tct att ata ttt tat ttt	3120
Thr Trp Ile Asn His Leu Val Thr Trp Gly Ser Ile Ile Phe Tyr Phe	
1025 1030 1035 1040	
gta ttt tcc ttg ttt tat gga ggg att ctg tgg cca ttt ttg ggc tcc	3168
Val Phe Ser Leu Phe Tyr Gly Gly Ile Leu Trp Pro Phe Leu Gly Ser	
1045 1050 1055	
cag aat atg tat ttt gtg ttt att cag ctg ctg tca agt ggt tct gct	3216
Gln Asn Met Tyr Phe Val Phe Ile Gln Leu Leu Ser Ser Gly Ser Ala	
1060 1065 1070	
tgg ttt gcc ata atc ctg atg gtt gtt aca tgt cta ttt ctt gat atc	3264
Trp Phe Ala Ile Ile Leu Met Val Val Thr Cys Leu Phe Leu Asp Ile	
1075 1080 1085	
ata aag aag gtc ttt gac cga cac ctg cac cct aca agt act gaa aag	3312
Ile Lys Lys Val Phe Asp Arg His Leu His Pro Thr Ser Thr Glu Lys	
1090 1095 1100	
gca cag ctt act gaa aca aat gca ggt atc aag tgc ttg gac tcc atg	3360
Ala Gln Leu Thr Glu Thr Asn Ala Gly Ile Lys Cys Leu Asp Ser Met	
1105 1110 1115 1120	
tgc tgt ttc ccg gaa gga gaa gca gog tgt gca tct gtt gga aga atg	3408
Cys Cys Phe Pro Glu Gly Glu Ala Ala Cys Ala Ser Val Gly Arg Met	
1125 1130 1135	
ctg gaa cga gtt ata gga aga tgt agt cca acc cac atc agc agt tca	3456
Leu Glu Arg Val Ile Gly Arg Cys Ser Pro Thr His Ile Ser Ser Ser	
1140 1145 1150	
tgg agt gca tcg gat cct ttc tat acc aac gac agg agc atc ttg act	3504
Trp Ser Ala Ser Asp Pro Phe Tyr Thr Asn Asp Arg Ser Ile Leu Thr	
1155 1160 1165	
ctc tcc aca atg gac tca tct act tgt taa	3534
Leu Ser Thr Met Asp Ser Ser Thr Cys *	
1170 1175	

<210> 22

<211> 4198

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (156)... (3410)

<400> 22

ggagtcgacc cacgcgtccg cattgagaca atgcctccac aaatacttga tgcaaaattc 60
 agtaagacag cacttggtga atcaccatta tagtttctga caaattgttc tcaaaaaggt 120
 accagctgga ggatgagtct gcgcatttgg atgaa atg cca cta atg atg tct 173

Met Pro Leu Met Met Ser
 1 5

gaa gaa ggc ttt gag aat gag gaa agt gat tac cac acc tta cca cga 221
 Glu Glu Gly Phe Glu Asn Glu Glu Ser Asp Tyr His Thr Leu Pro Arg
 10 15 20

gcc agg ata atg caa agg aaa aga gga ctg gag tgg ttt gtc tgt gat 269
 Ala Arg Ile Met Gln Arg Lys Arg Gly Leu Glu Trp Phe Val Cys Asp
 25 30 35

ggc tgg aag ttc ctc tgt acc agt tgc tgt ggt tgg ctg ata aat att 317
 Gly Trp Lys Phe Leu Cys Thr Ser Cys Cys Gly Trp Leu Ile Asn Ile
 40 45 50

tgt cga aga aag aaa gag ctg aaa gct cgc aca gta tgg ctt gga tgt 365
 Cys Arg Arg Lys Lys Glu Leu Lys Ala Arg Thr Val Trp Leu Gly Cys
 55 60 65 70

cct gaa aag tgt gaa gaa aaa cat ccc agg aat tct ata aaa aat caa 413
 Pro Glu Lys Cys Glu Glu Lys His Pro Arg Asn Ser Ile Lys Asn Gln
 75 80 85

aaa tac aat gtg ttt acc ttt ata cct ggg gtt ttg tat gaa caa ttc 461
 Lys Tyr Asn Val Phe Thr Phe Ile Pro Gly Val Leu Tyr Glu Gln Phe
 90 95 100

aag ttt ttc ttg aat ctc tat ttt cta gtg ata tcc tgc tca cag ttt 509
 Lys Phe Phe Leu Asn Leu Tyr Phe Leu Val Ile Ser Cys Ser Gln Phe
 105 110 115

gta cca gca ttg aaa ata ggc tat ctc tac acc tac tgg gct cct ctg 557
 Val Pro Ala Leu Lys Ile Gly Tyr Leu Tyr Thr Tyr Trp Ala Pro Leu
 120 125 130

gga ttt gtc ttg gct gtt act atg aca cgg gaa gca att gat gaa ttt 605
 Gly Phe Val Leu Ala Val Thr Met Thr Arg Glu Ala Ile Asp Glu Phe
 135 140 145 150

cgg cgt ttt cag cgt gac aag gaa gtg aat tca caa cta tat agc aag 653
 Arg Arg Phe Gln Arg Asp Lys Glu Val Asn Ser Gln Leu Tyr Ser Lys
 155 160 165

ctt aca gta aga ggt aaa gtg caa gtt aag agt tca gac ata caa gtt 701
 Leu Thr Val Arg Gly Lys Val Gln Val Lys Ser Ser Asp Ile Gln Val
 170 175 180

gga gac ctc atc ata gtg gaa aag aat caa aga att cca tcg gac atg	749
Gly Asp Leu Ile Ile Val Glu Lys Asn Gln Arg Ile Pro Ser Asp Met	
185 190 195	
gtg ttt ctt agg act tca gaa aaa gca ggt tcg tgt ttt att cga act	797
Val Phe Leu Arg Thr Ser Glu Lys Ala Gly Ser Cys Phe Ile Arg Thr	
200 205 210	
gat caa cta gat ggt gaa act gac tgg aag ctg aag gtg gca gtg agc	845
Asp Gln Leu Asp Gly Glu Thr Asp Trp Lys Leu Lys Val Ala Val Ser	
215 220 225 230	
tgc acg caa cag ctg ccg gct ctg ggg gac ctt ttt tct atc agt gct	893
Cys Thr Gln Gln Leu Pro Ala Leu Gly Asp Leu Phe Ser Ile Ser Ala	
235 240 245	
tat gtt tat gct cag aaa cca caa atg gac att cac agt ttc gaa ggc	941
Tyr Val Tyr Ala Gln Lys Pro Gln Met Asp Ile His Ser Phe Glu Gly	
250 255 260	
aca ttt acc agg gaa gac agt gac ccg ccc att cat gaa agt ctc agc	989
Thr Phe Thr Arg Glu Asp Ser Asp Pro Pro Ile His Glu Ser Leu Ser	
265 270 275	
ata gaa aat aca ttg tgg gca agc acc att gtt gca tca ggt act gta	1037
Ile Glu Asn Thr Leu Trp Ala Ser Thr Ile Val Ala Ser Gly Thr Val	
280 285 290	
ata ggt gtt gtc att tat acc gga aaa gag act cga agt gta atg aac	1085
Ile Gly Val Val Ile Tyr Thr Gly Lys Glu Thr Arg Ser Val Met Asn	
295 300 305 310	
aca tcc aat cca aaa aat aag gtt ggt ttg ttg gac ctt gaa ctc aat	1133
Thr Ser Asn Pro Lys Asn Lys Val Gly Leu Leu Asp Leu Glu Leu Asn	
315 320 325	
cgg ctg acg aaa gcg cta ttt ttg gct tta gtt gct ctt tcc att gtt	1181
Arg Leu Thr Lys Ala Leu Phe Leu Ala Leu Val Ala Leu Ser Ile Val	
330 335 340	
atg gta acc tta caa gga ttt gtg ggt cca tgg tac cgc aat ctt ttt	1229
Met Val Thr Leu Gln Gly Phe Val Gly Pro Trp Tyr Arg Asn Leu Phe	
345 350 355	
cgg ttc ctt ctc ctc ttt tct tac atc att ccc ata agt ttg cgt gtg	1277
Arg Phe Leu Leu Leu Phe Ser Tyr Ile Ile Pro Ile Ser Leu Arg Val	
360 365 370	
aac ttg gac atg ggc aaa gcg gtg tat gga tgg atg atg atg aaa gat	1325
Asn Leu Asp Met Gly Lys Ala Val Tyr Gly Trp Met Met Met Lys Asp	
375 380 385 390	
gag aac atc cct ggc acg gtc gtt cgg acc agc act atc cca gag gaa	1373
Glu Asn Ile Pro Gly Thr Val Val Arg Thr Ser Thr Ile Pro Glu Glu	
395 400 405	
ctt ggg cgc ctg gtg tat tta ttg aca gac aaa aca gga acc ctc acc	1421
Leu Gly Arg Leu Val Tyr Leu Leu Thr Asp Lys Thr Gly Thr Leu Thr	
410 415 420	

cag aat gaa atg ata ttt aag cgg ctg cac ctg ggc acc gtg tcc tat	1469
Gln Asn Glu Met Ile, Phe Lys, Arg Leu His Leu Gly Thr Val Ser Tyr	
425 430 435	
ggc gcc gac acg atg gat gag atc cag agc cat gtc agg gac tcc tac	1517
Gly Ala Asp Thr Met Asp Glu Ile Gln Ser His Val Arg Asp Ser Tyr	
440 445 450	
tca cag atg cag tct caa gct ggt gga aac aat act ggt tca act cca	1565
Ser Gln Met Gln Ser Gln Ala Gly Gly Asn Asn Thr Gly Ser Thr Pro	
455 460 465 470	
cta aga aaa gcc caa tct tca gct ccc aaa gtt agg aaa agt gtc agt	1613
Leu Arg Lys Ala Gln Ser Ser Ala Pro Lys Val Arg Lys Ser Val Ser	
475 480 485	
agt cga atc cat gaa gcc gtg aaa gcc atc gtg ctg tgt cac aac gtg	1661
Ser Arg Ile His Glu Ala Val Lys Ala Ile Val Leu Cys His Asn Val	
490 495 500	
acc ccc gtg tat gag tct cgg gcc ggc gtt act gag gag act gag ttc	1709
Thr Pro Val Tyr Glu Ser Arg Ala Gly Val Thr Glu Glu Thr Glu Phe	
505 510 515	
gca gag gct gac caa gac ttc agt gat gag aat cgc acc tac cag gct	1757
Ala Glu Ala Asp Gln Asp Phe Ser Asp Glu Asn Arg Thr Tyr Gln Ala	
520 525 530	
tec agc ccg gat gag gtc gct ctg gtg cag tgg aca gag agt gtg ggc	1805
Ser Ser Pro Asp Glu Val Ala Leu Val Gln Trp Thr Glu Ser Val Gly	
535 540 545 550	
ctc acg ctg gtc agc agg gac ctc acc tcc atg cag ctg aag acc ccc	1853
Leu Thr Leu Val Ser Arg Asp Leu Thr Ser Met Gln Leu Lys Thr Pro	
555 560 565	
agt ggc cag gtc ctc agc ttc tgc att ctg cag ctg ttt ccc ttc acc	1901
Ser Gly Gln Val Leu Ser Phe Cys Ile Leu Gln Leu Phe Pro Phe Thr	
570 575 580	
tcc gag agc aag cgg atg ggc gtc atc gtc agg gat gaa tcc acg gca	1949
Ser Glu Ser Lys Arg Met Gly Val Ile Val Arg Asp Glu Ser Thr Ala	
585 590 595	
gaa atc aca ttc tac atg aag ggc got gac gtg gcc atg tct cct atc	1997
Glu Ile Thr Phe Tyr Met Lys Gly Ala Asp Val Ala Met Ser Pro Ile	
600 605 610	
gtg cag tat aat gac tgg ctg gaa gag gag tgc gga aac atg gct cgc	2045
Val Gln Tyr Asn Asp Trp Leu Glu Glu Glu Cys Gly Asn Met Ala Arg	
615 620 625 630	
gaa gga ctg cgg acc ctc gtg gtt gca aag aag gcg ttg aca gag gag	2093
Glu Gly Leu Arg Thr Leu Val Val Ala Lys Lys Ala Leu Thr Glu Glu	
635 640 645	
cag tac cag gac ttt gag agc cga tac act caa gcc aag ctg agc atg	2141
Gln Tyr Gln Asp Phe Glu Ser Arg Tyr Thr Gln Ala Lys Leu Ser Met	
650 655 660	

cac gac agg tcc ctc aag gtg gcc gcg gta gtc gag agc ctg gag agg	2189
His Asp Arg Ser Leu Lys Val Ala Ala Val Val Glu Ser Leu Glu Arg	
665 670 675	
gag atg gaa ctg ctg tgc ctc acc ggc gtg gag gac cag ctg cag gca	2237
Glu Met Glu Leu Leu Cys Leu Thr Gly Val Glu Asp Gln Leu Gln Ala	
680 685 690	
gac gtg cgg ccc acg ctg gag atg ctg cgc aac gcc ggg atc aag ata	2285
Asp Val Arg Pro Thr Leu Glu Met Leu Arg Asn Ala Gly Ile Lys Ile	
695 700 705 710	
tgg atg cta aca ggc gat aaa ctc gag aca gct acc tgc att gcc aaa	2333
Trp Met Leu Thr Gly Asp Lys Leu Glu Thr Ala Thr Cys Ile Ala Lys	
715 720 725	
agt tca cat ctc gtg tct aga aca caa gat att cat att ttc aga cag	2381
Ser Ser His Leu Val Ser Arg Thr Gln Asp Ile His Ile Phe Arg Gln	
730 735 740	
gta acc agt cgg gga gag gca cat ttg gag ctg aat gca ttt cga agg	2429
Val Thr Ser Arg Gly Glu Ala His Leu Glu Leu Asn Ala Phe Arg Arg	
745 750 755	
aag cat gat tgt gca cta gtc ata tct ggg gac tct ctg gag gtt tgt	2477
Lys His Asp Cys Ala Leu Val Ile Ser Gly Asp Ser Leu Glu Val Cys	
760 765 770	
cta aag tac tac gag cat gaa ttt gtg gag ctg gcc tgc cag tgc cct	2525
Leu Lys Tyr Tyr Glu His Glu Phe Val Glu Leu Ala Cys Gln Cys Pro	
775 780 785 790	
gcc gtg gtt tgc tgc cgc tgc tca ccc acc cag aag gcc cgc att gtg	2573
Ala Val Val Cys Cys Arg Cys Ser Pro Thr Gln Lys Ala Arg Ile Val	
795 800 805	
aca ctg ctg cag cag cac aca ggg aga cgc acc tgc gcc atc ggt gat	2621
Thr Leu Leu Gln Gln His Thr Gly Arg Arg Thr Cys Ala Ile Gly Asp	
810 815 820	
gga gga aat gat gtc agc atg att cag gca gca gac tgt ggg att ggg	2669
Gly Gly Asn Asp Val Ser Met Ile Gln Ala Ala Asp Cys Gly Ile Gly	
825 830 835	
att gag gga aag gag ggt aaa cag gcc tgc ctg gcg gcc gac ttc tcc	2717
Ile Glu Gly Lys Glu Gly Lys Gln Ala Ser Leu Ala Ala Asp Phe Ser	
840 845 850	
atc acg cag ttc cgg cac ata ggc agg ctg ctc atg gtg cac ggg cgg	2765
Ile Thr Gln Phe Arg His Ile Gly Arg Leu Leu Met Val His Gly Arg	
855 860 865 870	
aac agc tac aag agg tgc gcg gca ctc ggc cag ttc gtc atg cac agg	2813
Asn Ser Tyr Lys Arg Ser Ala Ala Leu Gly Gln Phe Val Met His Arg	
875 880 885	
ggc ctt atc atc tcc acc atg cag gct gtg ttt tcc tca gtc ttc tac	2861
Gly Leu Ile Ile Ser Thr Met Gln Ala Val Phe Ser Ser Val Phe Tyr	
890 895 900	

ttc gca tcc gtc cct ttg tat cag ggc ttc ctc atg gtg ggg tat gcc	2909
Phe Ala Ser Val Pro Leu Tyr Gln Gly Phe Leu Met Val Gly Tyr Ala	
905 910 915	
acc ata tac acc atg ttc cca gtg ttc tcc tta gtg ctg gac cag gac	2957
Thr Ile Tyr Thr Met Phe Pro Val Phe Ser Leu Val Leu Asp Gln Asp	
920 925 930	
gtg aag cca gag atg gcg atg ctc tac ccg gag ctg tac aag gac ctc	3005
Val Lys Pro Glu Met Ala Met Leu Tyr Pro Glu Leu Tyr Lys Asp Leu	
935 940 945 950	
acc aag gga aga tcc ttg tcc ttc aaa acc ttc ctc atc tgg gtt tta	3053
Thr Lys Gly Arg Ser Leu Ser Phe Lys Thr Phe Leu Ile Trp Val Leu	
955 960 965	
ata agt att tac caa ggc ggc atc ctc atg tat ggg gcc ctg gtg ctc	3101
Ile Ser Ile Tyr Gln Gly Gly Ile Leu Met Tyr Gly Ala Leu Val Leu	
970 975 980	
ttc gag tct gag ttc gtc cac gtg gtg gcc atc tcc ttc acc gca ctg	3149
Phe Glu Ser Glu Phe Val His Val Val Ala Ile Ser Phe Thr Ala Leu	
985 990 995	
atc ctg acc gag ctg ctg atg gtg gcg ctg acc gtc cgc acg tgg cac	3197
Ile Leu Thr Glu Leu Leu Met Val Ala Leu Thr Val Arg Thr Trp His	
1000 1005 1010	
tgg ctg atg gtg gtg gcc gag ttc ctc agc tta ggc tgc tac gtg tcc	3245
Trp Leu Met Val Val Ala Glu Phe Leu Ser Leu Gly Cys Tyr Val Ser	
1015 1020 1025 1030	
tca ctc gct ttt ctc aat gaa tat ttt gat gtt gcc ttt atc acc acc	3293
Ser Leu Ala Phe Leu Asn Glu Tyr Phe Asp Val Ala Phe Ile Thr Thr	
1035 1040 1045	
gtg acc ttc ctg tgg aaa gtg tgc gcg atc acc gtg gtc agc tgc ctc	3341
Val Thr Phe Leu Trp Lys Val Ser Ala Ile Thr Val Val Ser Cys Leu	
1050 1055 1060	
ccg ctg tat gtc ctc aag tac ctg agg cgc aag tct tct cct ccc agc	3389
Pro Leu Tyr Val Leu Lys Tyr Leu Arg Arg Lys Ser Ser Pro Pro Ser	
1065 1070 1075	
tac tgc aag ctg gcc tcc taa ggggctgtgc acccccagcg ggctggcccc	3440
Tyr Cys Lys Leu Ala Ser *	
1080	
agcaccttct gcccttccca gcacctgtg cccttgccag tgaacgcagg gtttgccatt	3500
gctaccaagc aagcaccaca agaaaggagg ggtacgccag gcgagcccag ggcacagatg	3560
ctgagacagc ctctccttct cagtgcaggg acgtcacccc tgccaggcaa gcccagggca	3620
cagatgccag gatggcttct ccctctcagt gcgaggcttc acccctgcca ggcaagccca	3680
gggcatagat gctgagacag cctctccctc tcagtgcagg gacgtcaccc ctgccaggca	3740
agcccagggc acagaggccg ggacggcctc tccctctcag tgtgaggctt caccatgct	3800
aggcaagccc agggcacaga tgccgggatg gccctccct ctcagtgcgg gaacgtcacc	3860
cctgccaggc aagcccaggg cacagatgct gcgatggcct ctccctctta agtggtgggc	3920
ctcacccctg cttttctttc tttttttgta ttgtcaaaat tgtatttcca tattgaagca	3980
gcttgagttt ctactgaaaa tgagcccgaa ttatttctact attactgtaa agggttcatc	4040
ttactctggc attctgagaa ttagactgaa agtttaatt ctgcagttcc ctcatattca	4100
gattctttct ttgatgttat aacacaaagt ctttctact caaatgtaat aaaattgagg	4160
ctccacggag aaaaaaaaaa aaaaaaaaaa aaaaaaaa	4198

<210> 23

<211> 1084

<212> PRT

<213> Homo sapiens

<400> 23

```

Met Pro Leu Met Met Ser Glu Glu Gly Phe Glu Asn Glu Glu Ser Asp
1          5          10          15
Tyr His Thr Leu Pro Arg Ala Arg Ile Met Gln Arg Lys Arg Gly Leu
20          25          30
Glu Trp Phe Val Cys Asp Gly Trp Lys Phe Leu Cys Thr Ser Cys Cys
35          40          45
Gly Trp Leu Ile Asn Ile Cys Arg Arg Lys Lys Glu Leu Lys Ala Arg
50          55          60
Thr Val Trp Leu Gly Cys Pro Glu Lys Cys Glu Glu Lys His Pro Arg
65          70          75          80
Asn Ser Ile Lys Asn Gln Lys Tyr Asn Val Phe Thr Phe Ile Pro Gly
85          90          95
Val Leu Tyr Glu Gln Phe Lys Phe Phe Leu Asn Leu Tyr Phe Leu Val
100          105          110
Ile Ser Cys Ser Gln Phe Val Pro Ala Leu Lys Ile Gly Tyr Leu Tyr
115          120          125
Thr Tyr Trp Ala Pro Leu Gly Phe Val Leu Ala Val Thr Met Thr Arg
130          135          140
Glu Ala Ile Asp Glu Phe Arg Arg Phe Gln Arg Asp Lys Glu Val Asn
145          150          155          160
Ser Gln Leu Tyr Ser Lys Leu Thr Val Arg Gly Lys Val Gln Val Lys
165          170          175
Ser Ser Asp Ile Gln Val Gly Asp Leu Ile Ile Val Glu Lys Asn Gln
180          185          190
Arg Ile Pro Ser Asp Met Val Phe Leu Arg Thr Ser Glu Lys Ala Gly
195          200          205
Ser Cys Phe Ile Arg Thr Asp Gln Leu Asp Gly Glu Thr Asp Trp Lys
210          215          220
Leu Lys Val Ala Val Ser Cys Thr Gln Gln Leu Pro Ala Leu Gly Asp
225          230          235          240
Leu Phe Ser Ile Ser Ala Tyr Val Tyr Ala Gln Lys Pro Gln Met Asp
245          250          255
Ile His Ser Phe Glu Gly Thr Phe Thr Arg Glu Asp Ser Asp Pro Pro
260          265          270
Ile His Glu Ser Leu Ser Ile Glu Asn Thr Leu Trp Ala Ser Thr Ile
275          280          285
Val Ala Ser Gly Thr Val Ile Gly Val Val Ile Tyr Thr Gly Lys Glu
290          295          300
Thr Arg Ser Val Met Asn Thr Ser Asn Pro Lys Asn Lys Val Gly Leu
305          310          315          320
Leu Asp Leu Glu Leu Asn Arg Leu Thr Lys Ala Leu Phe Leu Ala Leu
325          330          335
Val Ala Leu Ser Ile Val Met Val Thr Leu Gln Gly Phe Val Gly Pro
340          345          350
Trp Tyr Arg Asn Leu Phe Arg Phe Leu Leu Leu Phe Ser Tyr Ile Ile
355          360          365
Pro Ile Ser Leu Arg Val Asn Leu Asp Met Gly Lys Ala Val Tyr Gly
370          375          380
Trp Met Met Met Lys Asp Glu Asn Ile Pro Gly Thr Val Val Arg Thr
385          390          395          400
Ser Thr Ile Pro Glu Leu Gly Arg Leu Val Tyr Leu Leu Thr Asp
405          410          415
Lys Thr Gly Thr Leu Thr Gln Asn Glu Met Ile Phe Lys Arg Leu His
420          425          430

```

- 84 -

Leu Met Val Gly Tyr Ala Thr Ile Tyr Thr Met Phe Pro Val Phe Ser
 915 920 925
 Leu Val Leu Asp Gln Asp Val Lys Pro Glu Met Ala Met Leu Tyr Pro
 930 935 940
 Glu Leu Tyr Lys Asp Leu Thr Lys Gly Arg Ser Leu Ser Phe Lys Thr
 945 950 955 960
 Phe Leu Ile Trp Val Leu Ile Ser Ile Tyr Gln Gly Gly Ile Leu Met
 965 970 975
 Tyr Gly Ala Leu Val Leu Phe Glu Ser Glu Phe Val His Val Val Ala
 980 985 990
 Ile Ser Phe Thr Ala Leu Ile Leu Thr Glu Leu Leu Met Val Ala Leu
 995 1000 1005
 Thr Val Arg Thr Trp His Trp Leu Met Val Val Ala Glu Phe Leu Ser
 1010 1015 1020
 Leu Gly Cys Tyr Val Ser Ser Leu Ala Phe Leu Asn Glu Tyr Phe Asp
 1025 1030 1035 1040
 Val Ala Phe Ile Thr Thr Val Thr Phe Leu Trp Lys Val Ser Ala Ile
 1045 1050 1055
 Thr Val Val Ser Cys Leu Pro Leu Tyr Val Leu Lys Tyr Leu Arg Arg
 1060 1065 1070
 Lys Ser Ser Pro Pro Ser Tyr Cys Lys Leu Ala Ser
 1075 1080

<210> 24

<211> 3255

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(3255)

<400> 24

atg cca cta atg atg tct gaa gaa ggc ttt gag aat gag gaa agt gat 48
 Met Pro Leu Met Met Ser Glu Glu Gly Phe Glu Asn Glu Glu Ser Asp
 1 5 10 15
 tac cac acc tta cca cga gcc agg ata atg caa agg aaa aga gga ctg 96
 Tyr His Thr Leu Pro Arg Ala Arg Ile Met Gln Arg Lys Arg Gly Leu
 20 25 30
 gag tgg ttt gtc tgt gat ggc tgg aag ttc ctc tgt acc agt tgc tgt 144
 Glu Trp Phe Val Cys Asp Gly Trp Lys Phe Leu Cys Thr Ser Cys Cys
 35 40 45
 ggt tgg ctg ata aat att tgt cga aga aag aaa gag ctg aaa gct cgc 192
 Gly Trp Leu Ile Asn Ile Cys Arg Arg Lys Lys Glu Leu Lys Ala Arg
 50 55 60
 aca gta tgg ctt gga tgt cct gaa aag tgt gaa gaa aaa cat ccc agg 240
 Thr Val Trp Leu Gly Cys Pro Glu Lys Cys Glu Glu Lys His Pro Arg
 65 70 75 80
 aat tct ata aaa aat caa aaa tac aat gtg ttt acc ttt ata cct ggg 288
 Asn Ser Ile Lys Asn Gln Lys Tyr Asn Val Phe Thr Phe Ile Pro Gly
 85 90 95
 gtt ttg tat gaa caa ttc aag ttt ttc ttg aat ctc tat ttt cta gtg 336
 Val Leu Tyr Glu Gln Phe Lys Phe Phe Leu Asn Leu Tyr Phe Leu Val
 100 105 110

ata tcc tgc tca cag ttt gta cca gca ttg aaa ata ggc tat ctc tac	384
Ile Ser Cys Ser Gln Phe Val Pro Ala Leu Lys Ile Gly Tyr Leu Tyr	
115 120 125	
acc tac tgg gct cct ctg gga ttt gtc ttg gct gtt act atg aca cgg	432
Thr Tyr Trp Ala Pro Leu Gly Phe Val Leu Ala Val Thr Met Thr Arg	
130 135 140	
gaa gca att gat gaa ttt cgg cgt ttt cag cgt gac aag gaa gtg aat	480
Glu Ala Ile Asp Glu Phe Arg Arg Phe Gln Arg Asp Lys Glu Val Asn	
145 150 155 160	
tca caa cta tat agc aag ctt aca gta aga ggt aaa gtg caa gtt aag	528
Ser Gln Leu Tyr Ser Lys Leu Thr Val Arg Gly Lys Val Gln Val Lys	
165 170 175	
agt tca gac ata caa gtt gga gac ctc atc ata gtg gaa aag aat caa	576
Ser Ser Asp Ile Gln Val Gly Asp Leu Ile Ile Val Glu Lys Asn Gln	
180 185 190	
aga att cca tcg gac atg gtg ttt ctt agg act tca gaa aaa gca ggt	624
Arg Ile Pro Ser Asp Met Val Phe Leu Arg Thr Ser Glu Lys Ala Gly	
195 200 205	
tcg tgt ttt att cga act gat caa cta gat ggt gaa act gac tgg aag	672
Ser Cys Phe Ile Arg Thr Asp Gln Leu Asp Gly Glu Thr Asp Trp Lys	
210 215 220	
ctg aag gtg gca gtg agc tgc acg caa cag ctg ccg gct ctg ggg gac	720
Leu Lys Val Ala Val Ser Cys Thr Gln Gln Leu Pro Ala Leu Gly Asp	
225 230 235 240	
ctt ttt tct atc agt gct tat gtt tat gct cag aaa cca caa atg gac	768
Leu Phe Ser Ile Ser Ala Tyr Val Tyr Ala Gln Lys Pro Gln Met Asp	
245 250 255	
att cac agt ttc gaa ggc aca ttt acc agg gaa gac agt gac ccg ccc	816
Ile His Ser Phe Glu Gly Thr Phe Thr Arg Glu Asp Ser Asp Pro Pro	
260 265 270	
att cat gaa agt ctc agc ata gaa aat aca ttg tgg gca agc acc att	864
Ile His Glu Ser Leu Ser Ile Glu Asn Thr Leu Trp Ala Ser Thr Ile	
275 280 285	
gtt gca tca ggt act gta ata ggt gtt gtc att tat acc gga aaa gag	912
Val Ala Ser Gly Thr Val Ile Gly Val Val Ile Tyr Thr Gly Lys Glu	
290 295 300	
act cga agt gta atg aac aca tcc aat cca aaa aat aag gtt ggt ttg	960
Thr Arg Ser Val Met Asn Thr Ser Asn Pro Lys Asn Lys Val Gly Leu	
305 310 315 320	
ttg gac ctt gaa ctc aat cgg ctg acg aaa gcg cta ttt ttg gct tta	1008
Leu Asp Leu Glu Leu Asn Arg Leu Thr Lys Ala Leu Phe Leu Ala Leu	
325 330 335	
gtt gct ctt tcc att gtt atg gta acc tta caa gga ttt gtg ggt cca	1056
Val Ala Leu Ser Ile Val Met Val Thr Leu Gln Gly Phe Val Gly Pro	
340 345 350	

tgg tac cgc aat ctt ttt cgg ttc ctt ctc ctc ttt tct tac atc att	1104
Trp Tyr Arg Asn Leu Phe Arg Phe Leu Leu Leu Phe Ser Tyr Ile Ile	
355 360 365	
ccc ata agt ttg cgt gtg aac ttg gac atg ggc aaa gcg gtg tat gga	1152
Pro Ile Ser Leu Arg Val Asn Leu Asp Met Gly Lys Ala Val Tyr Gly	
370 375 380	
tgg atg atg atg aaa gat gag aac atc cct ggc acg gtc gtt cgg acc	1200
Trp Met Met Met Lys Asp Glu Asn Ile Pro Gly Thr Val Val Arg Thr	
385 390 395 400	
agc act atc cca gag gaa ctt ggg cgc ctg gtg tat tta ttg aca gac	1248
Ser Thr Ile Pro Glu Glu Leu Gly Arg Leu Val Tyr Leu Leu Thr Asp	
405 410 415	
aaa aca gga acc ctc acc cag aat gaa atg ata ttt aag cgg ctg cac	1296
Lys Thr Gly Thr Leu Thr Gln Asn Glu Met Ile Phe Lys Arg Leu His	
420 425 430	
ctg ggc acc gtg tcc tat ggc gcc gac acg atg gat gag atc cag agc	1344
Leu Gly Thr Val Ser Tyr Gly Ala Asp Thr Met Asp Glu Ile Gln Ser	
435 440 445	
cat gtc agg gac tcc tac tca cag atg cag tct caa gct ggt gga aac	1392
His Val Arg Asp Ser Tyr Ser Gln Met Gln Ser Gln Ala Gly Gly Asn	
450 455 460	
aat act ggt tca act cca cta aga aaa gcc caa tct tca gct ccc aaa	1440
Asn Thr Gly Ser Thr Pro Leu Arg Lys Ala Gln Ser Ser Ala Pro Lys	
465 470 475 480	
gtt agg aaa agt gtc agt agt cga atc cat gaa gcc gtg aaa gcc atc	1488
Val Arg Lys Ser Val Ser Ser Arg Ile His Glu Ala Val Lys Ala Ile	
485 490 495	
gtg ctg tgt cac aac gtg acc ccc gtg tat gag tct cgg gcc ggc gtt	1536
Val Leu Cys His Asn Val Thr Pro Val Tyr Glu Ser Arg Ala Gly Val	
500 505 510	
act gag gag act gag ttc gca gag gct gac caa gac ttc agt gat gag	1584
Thr Glu Glu Thr Glu Phe Ala Glu Ala Asp Gln Asp Phe Ser Asp Glu	
515 520 525	
aat cgc acc tac cag gct tcc agc ccg gat gag gtc gct ctg gtg cag	1632
Asn Arg Thr Tyr Gln Ala Ser Ser Pro Asp Glu Val Ala Leu Val Gln	
530 535 540	
tgg aca gag agt gtg ggc ctc acg ctg gtc agc agg gac ctc acc tcc	1680
Trp Thr Glu Ser Val Gly Leu Thr Leu Val Ser Arg Asp Leu Thr Ser	
545 550 555 560	
atg cag ctg aag acc ccc agt ggc cag gtc ctc agc ttc tgc att ctg	1728
Met Gln Leu Lys Thr Pro Ser Gly Gln Val Leu Ser Phe Cys Ile Leu	
565 570 575	
cag ctg ttt ccc ttc acc tcc gag agc aag cgg atg ggc gtc atc gtc	1776
Gln Leu Phe Pro Phe Thr Ser Glu Ser Lys Arg Met Gly Val Ile Val	
580 585 590	

agg gat gaa tcc acg gca gaa atc aca ttc tac atg aag ggc gct gac	1824
Arg Asp Glu Ser Thr Ala Glu Ile Thr Phe Tyr Met Lys Gly Ala Asp	
595 600 605	
gtg gcc atg tct cct atc gtg cag tat aat gac tgg ctg gaa gag gag	1872
Val Ala Met Ser Pro Ile Val Gln Tyr Asn Asp Trp Leu Glu Glu Glu	
610 615 620	
tgc gga aac atg gct cgc gaa gga ctg cgg acc ctc gtg gtt gca aag	1920
Cys Gly Asn Met Ala Arg Glu Gly Leu Arg Thr Leu Val Val Ala Lys	
625 630 635 640	
aag gcg ttg aca gag gag cag tac cag gac ttt gag agc cga tac act	1968
Lys Ala Leu Thr Glu Glu Gln Tyr Gln Asp Phe Glu Ser Arg Tyr Thr	
645 650 655	
caa gcc aag ctg agc atg cac gac agg tcc ctc aag gtg gcc gcg gta	2016
Gln Ala Lys Leu Ser Met His Asp Arg Ser Leu Lys Val Ala Ala Val	
660 665 670	
gtc gag agc ctg gag agg gag atg gaa ctg ctg tgc ctc acc ggc gtg	2064
Val Glu Ser Leu Glu Arg Glu Met Glu Leu Leu Cys Leu Thr Gly Val	
675 680 685	
gag gac cag ctg cag gca gac gtg cgg ccc acg ctg gag atg ctg cgc	2112
Glu Asp Gln Leu Gln Ala Asp Val Arg Pro Thr Leu Glu Met Leu Arg	
690 695 700	
aac gcc ggg atc aag ata tgg atg cta aca ggc gat aaa ctc gag aca	2160
Asn Ala Gly Ile Lys Ile Trp Met Leu Thr Gly Asp Lys Leu Glu Thr	
705 710 715 720	
gct acc tgc att gcc aaa agt tca cat ctc gtg tct aga aca caa gat	2208
Ala Thr Cys Ile Ala Lys Ser Ser His Leu Val Ser Arg Thr Gln Asp	
725 730 735	
att cat att ttc aga cag gta acc agt cgg gga gag gca cat ttg gag	2256
Ile His Ile Phe Arg Gln Val Thr Ser Arg Gly Glu Ala His Leu Glu	
740 745 750	
ctg aat gca ttt cga agg aag cat gat tgt gca cta gtc ata tct ggg	2304
Leu Asn Ala Phe Arg Arg Lys His Asp Cys Ala Leu Val Ile Ser Gly	
755 760 765	
gac tct ctg gag gtt tgt cta aag tac tac gag cat gaa ttt gtg gag	2352
Asp Ser Leu Glu Val Cys Leu Lys Tyr Tyr Glu His Glu Phe Val Glu	
770 775 780	
ctg gcc tgc cag tgc cct gcc gtg gtt tgc tgc cgc tgc tca ccc acc	2400
Leu Ala Cys Gln Cys Pro Ala Val Val Cys Cys Arg Cys Ser Pro Thr	
785 790 795 800	
cag aag gcc cgc att gtg aca ctg ctg cag cag cac aca ggg aga cgc	2448
Gln Lys Ala Arg Ile Val Thr Leu Leu Gln Gln His Thr Gly Arg Arg	
805 810 815	
acc tgc gcc atc ggt gat gga gga aat gat gtc agc atg att cag gca	2496
Thr Cys Ala Ile Gly Asp Gly Gly Asn Asp Val Ser Met Ile Gln Ala	
820 825 830	

gca gac tgt ggg att ggg att gag gga aag gag ggt aaa cag gcc tcg 2544
 Ala Asp Cys Gly Ile Gly Ile Glu Gly Lys Glu Gly Lys Gln Ala Ser
 835 840 845

ctg gcg gcc gac ttc tcc atc acg cag ttc cgg cac ata ggc agg ctg 2592
 Leu Ala Ala Asp Phe Ser Ile Thr Gln Phe Arg His Ile Gly Arg Leu
 850 855 860

ctc atg gtg cac ggg cgg aac agc tac aag agg tcg gcg gca ctc ggc 2640
 Leu Met Val His Gly Arg Asn Ser Tyr Lys Arg Ser Ala Ala Leu Gly
 865 870 875 880

cag ttc gtc atg cac agg ggc ctt atc atc tcc acc atg cag gct gtg 2688
 Gln Phe Val Met His Arg Gly Leu Ile Ile Ser Thr Met Gln Ala Val
 885 890 895

ttt tcc tca gtc ttc tac ttc gca tcc gtc cct ttg tat cag ggc ttc 2736
 Phe Ser Ser Val Phe Tyr Phe Ala Ser Val Pro Leu Tyr Gln Gly Phe
 900 905 910

ctc atg gtg ggg tat gcc acc ata tac acc atg ttc cca gtg ttc tcc 2784
 Leu Met Val Gly Tyr Ala Thr Ile Tyr Thr Met Phe Pro Val Phe Ser
 915 920 925

tta gtg ctg gac cag gac gtg aag cca gag atg gcg atg ctc tac ccg 2832
 Leu Val Leu Asp Gln Asp Val Lys Pro Glu Met Ala Met Leu Tyr Pro
 930 935 940

gag ctg tac aag gac ctc acc aag gga aga tcc ttg tcc ttc aaa acc 2880
 Glu Leu Tyr Lys Asp Leu Thr Lys Gly Arg Ser Leu Ser Phe Lys Thr
 945 950 955 960

ttc ctc atc tgg gtt tta ata agt att tac caa ggc ggc atc ctc atg 2928
 Phe Leu Ile Trp Val Leu Ile Ser Ile Tyr Gln Gly Gly Ile Leu Met
 965 970 975

tat ggg gcc ctg gtg ctc ttc gag tct gag ttc gtc cac gtg gtg gcc 2976
 Tyr Gly Ala Leu Val Leu Phe Glu Ser Glu Phe Val His Val Val Ala
 980 985 990

atc tcc ttc acc gca ctg atc ctg acc gag ctg ctg atg gtg gcg ctg 3024
 Ile Ser Phe Thr Ala Leu Ile Leu Thr Glu Leu Leu Met Val Ala Leu
 995 1000 1005

acc gtc cgc acg tgg cac tgg ctg atg gtg gtg gcc gag ttc ctc agc 3072
 Thr Val Arg Thr Trp His Trp Leu Met Val Val Ala Glu Phe Leu Ser
 1010 1015 1020

tta ggc tgc tac gtg tcc tca ctc gct ttt ctc aat gaa tat ttt gat 3120
 Leu Gly Cys Tyr Val Ser Ser Leu Ala Phe Leu Asn Glu Tyr Phe Asp
 1025 1030 1035 1040

gtt gcc ttt atc acc acc gtg acc ttc ctg tgg aaa gtg tcg gcg atc 3168
 Val Ala Phe Ile Thr Thr Val Thr Phe Leu Trp Lys Val Ser Ala Ile
 1045 1050 1055

acc gtg gtc agc tgc ctc ccg ctg tat gtc ctc aag tac ctg agg cgc 3216
 Thr Val Val Ser Cys Leu Pro Leu Tyr Val Leu Lys Tyr Leu Arg Arg
 1060 1065 1070

aag tct tct cct ccc agc tac tgc aag ctg gcc tcc taa
 Lys Ser Ser Pro Pro Ser Tyr Cys Lys Leu Ala Ser *
 1075 1080

3255

<210> 25
 <211> 4231
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (156)...(3443)

<400> 25

ggagtcgacc cagcgcgtccg cattgagaca atgcctccac aaataacttga tgcaaaattc 60
 agtaagacag cacttggttga atcaccatta tagtttctga caaattgttc tcaaaaaggt 120
 accagctgga ggaatgagtct gcgcatttgg atgaa atg cca cta atg atg tct 173
 Met Pro Leu Met Met Ser

1 5

gaa gaa ggc ttt gag aat gag gaa agt gat tac cac acc tta cca cga 221
 Glu Glu Gly Phe Glu Asn Glu Glu Ser Asp Tyr His Thr Leu Pro Arg
 10 15 20

gcc agg ata atg caa agg aaa aga gga ctg gag tgg ttt gtc tgt gat 269
 Ala Arg Ile Met Gln Arg Lys Arg Gly Leu Glu Trp Phe Val Cys Asp
 25 30 35

ggc tgg aag ttc ctc tgt acc agt tgc tgt ggt tgg ctg ata aat att 317
 Gly Trp Lys Phe Leu Cys Thr Ser Cys Cys Gly Trp Leu Ile Asn Ile
 40 45 50

tgt cga aga aag aaa gag ctg aaa gct cgc aca gta tgg ctt gga tgt 365
 Cys Arg Arg Lys Lys Glu Leu Lys Ala Arg Thr Val Trp Leu Gly Cys
 55 60 65 70

cct gaa aag tgt gaa gaa aaa cat ccc agg aat tct ata aaa aat caa 413
 Pro Glu Lys Cys Glu Glu Lys His Pro Arg Asn Ser Ile Lys Asn Gln
 75 80 85

aaa tac aat gtg ttt acc ttt ata cct ggg gtt ttg tat gaa caa ttc 461
 Lys Tyr Asn Val Phe Thr Phe Ile Pro Gly Val Leu Tyr Glu Gln Phe
 90 95 100

aag ttt ttc ttg aat ctc tat ttt cta gtg ata tcc tgc tca cag ttt 509
 Lys Phe Phe Leu Asn Leu Tyr Phe Leu Val Ile Ser Cys Ser Gln Phe
 105 110 115

gta cca gca ttg aaa ata ggc tat ctc tac acc tac tgg gct cct ctg 557
 Val Pro Ala Leu Lys Ile Gly Tyr Leu Tyr Thr Tyr Trp Ala Pro Leu
 120 125 130

gga ttt gtc ttg gct gtt act atg aca cgg gaa gca att gat gaa ttt 605
 Gly Phe Val Leu Ala Val Thr Met Thr Arg Glu Ala Ile Asp Glu Phe
 135 140 145 150

cgg cgt ttt cag cgt gac aag gaa gtg aat tca caa cta tat agc aag 653
 Arg Arg Phe Gln Arg Asp Lys Glu Val Asn Ser Gln Leu Tyr Ser Lys
 155 160 165

ctt	aca	gta	aga	ggg	aaa	gtg	caa	gtt	aag	agt	tca	gac	ata	caa	gtt	701
Leu	Thr	Val	Arg	Gly	Lys	Val	Gln	Val	Lys	Ser	Ser	Asp	Ile	Gln	Val	
			170					175							180	
gga	gac	ctc	atc	ata	gtg	gaa	aag	aat	caa	aga	att	cca	tgc	gac	atg	749
Gly	Asp	Leu	Ile	Ile	Val	Glu	Lys	Asn	Gln	Arg	Ile	Pro	Ser	Asp	Met	
		185					190					195				
gtg	ttt	ctt	agg	act	tca	gaa	aaa	gca	ggg	tgc	tgt	ttt	att	cga	act	797
Val	Phe	Leu	Arg	Thr	Ser	Glu	Lys	Ala	Gly	Ser	Cys	Phe	Ile	Arg	Thr	
	200					205					210					
gat	caa	cta	gat	ggg	gaa	act	gac	tgg	aag	ctg	aag	gtg	gca	gtg	agc	845
Asp	Gln	Leu	Asp	Gly	Glu	Thr	Asp	Trp	Lys	Leu	Lys	Val	Ala	Val	Ser	
	215				220				225						230	
tgc	acg	caa	cag	ctg	ccg	gct	ctg	ggg	gac	ctt	ttt	tct	atc	agt	gct	893
Cys	Thr	Gln	Gln	Leu	Pro	Ala	Leu	Gly	Asp	Leu	Phe	Ser	Ile	Ser	Ala	
			235					240						245		
tat	gtt	tat	gct	cag	aaa	cca	caa	atg	gac	att	cac	agt	ttc	gaa	ggc	941
Tyr	Val	Tyr	Ala	Gln	Lys	Pro	Gln	Met	Asp	Ile	His	Ser	Phe	Glu	Gly	
			250					255						260		
aca	ttt	acc	agg	gaa	gac	agt	gac	ccg	ccc	att	cat	gaa	agt	ctc	agc	989
Thr	Phe	Thr	Arg	Glu	Asp	Ser	Asp	Pro	Pro	Ile	His	Glu	Ser	Leu	Ser	
		265					270					275				
ata	gaa	aat	aca	ttg	tgg	gca	agc	acc	att	gtt	gca	tca	ggg	act	gta	1037
Ile	Glu	Asn	Thr	Leu	Trp	Ala	Ser	Thr	Ile	Val	Ala	Ser	Gly	Thr	Val	
	280					285					290					
ata	ggg	gtt	gtc	att	tat	acc	gga	aaa	gag	act	cga	agt	gta	atg	aac	1085
Ile	Gly	Val	Val	Ile	Tyr	Thr	Gly	Lys	Glu	Thr	Arg	Ser	Val	Met	Asn	
	295				300					305					310	
aca	tcc	aat	cca	aaa	aat	aag	gtt	ggg	ttg	ttg	gac	ctt	gaa	ctc	aat	1133
Thr	Ser	Asn	Pro	Lys	Asn	Lys	Val	Gly	Leu	Leu	Asp	Leu	Glu	Leu	Asn	
			315					320					325			
cgg	ctg	acg	aaa	gcg	cta	ttt	ttg	gct	tta	gtt	gct	ctt	tcc	att	gtt	1181
Arg	Leu	Thr	Lys	Ala	Leu	Phe	Leu	Ala	Leu	Val	Ala	Leu	Ser	Ile	Val	
			330					335					340			
atg	gta	acc	tta	caa	gga	ttt	gtg	ggg	cca	tgg	tac	cgc	aat	ctt	ttt	1229
Met	Val	Thr	Leu	Gln	Gly	Phe	Val	Gly	Pro	Trp	Tyr	Arg	Asn	Leu	Phe	
		345					350					355				
cgg	ttc	ctt	ctc	ctc	ttt	tct	tac	atc	att	ccc	ata	agt	ttg	cgt	gtg	1277
Arg	Phe	Leu	Leu	Leu	Phe	Ser	Tyr	Ile	Ile	Pro	Ile	Ser	Leu	Arg	Val	
	360					365					370					
aac	ttg	gac	atg	ggc	aaa	gcg	gtg	tat	gga	tgg	atg	atg	atg	aaa	gat	1325
Asn	Leu	Asp	Met	Gly	Lys	Ala	Val	Tyr	Gly	Trp	Met	Met	Met	Lys	Asp	
	375				380					385					390	
gag	aac	atc	cct	ggc	acg	gtc	gtt	cgg	acc	agc	act	atc	cca	gag	gaa	1373
Glu	Asn	Ile	Pro	Gly	Thr	Val	Val	Arg	Thr	Ser	Thr	Ile	Pro	Glu	Glu	
			395					400						405		

ctt ggg cgc ctg gtg tat tta ttg aca gac aaa aca gga acc ctc acc	1421
Leu Gly Arg Leu Val Tyr Leu Leu Thr Asp Lys Thr Gly Thr Leu Thr	
410 --- 415 420	
cag aat gaa atg ata ttt aag cgg ctg cac ctg ggc acc gtg tcc tat	1469
Gln Asn Glu Met Ile Phe Lys Arg Leu His Leu Gly Thr Val Ser Tyr	
425 430 435	
ggc gcc gac acg atg gat gag atc cag agc cat gtc agg gac tcc tac	1517
Gly Ala Asp Thr Met Asp Glu Ile Gln Ser His Val Arg Asp Ser Tyr	
440 445 450	
tca cag atg cag tct caa gct ggt gga aac aat act ggt tca act cca	1565
Ser Gln Met Gln Ser Gln Ala Gly Gly Asn Asn Thr Gly Ser Thr Pro	
455 460 465 470	
cta aga aaa gcc caa tct tca gct ccc aaa gtt agg aaa agt gtc agt	1613
Leu Arg Lys Ala Gln Ser Ser Ala Pro Lys Val Arg Lys Ser Val Ser	
475 480 485	
agt cga atc cat gaa gcc gtg aaa gcc atc gtg ctg tgt cac aac gtg	1661
Ser Arg Ile His Glu Ala Val Lys Ala Ile Val Leu Cys His Asn Val	
490 495 500	
acc ccc gtg tat gag tct cgg gcc ggc gtt act gag gag act gag ttc	1709
Thr Pro Val Tyr Glu Ser Arg Ala Gly Val Thr Glu Glu Thr Glu Phe	
505 510 515	
gca gag gct gac caa gac ttc agt gat gag aat cgc acc tac cag gct	1757
Ala Glu Ala Asp Gln Asp Phe Ser Asp Glu Asn Arg Thr Tyr Gln Ala	
520 525 530	
tcc agc ccg gat gag gtc gct ctg gtg cag tgg aca gag agt gtg ggc	1805
Ser Ser Pro Asp Glu Val Ala Leu Val Gln Trp Thr Glu Ser Val Gly	
535 540 545 550	
ctc acg ctg gtc agc agg gac ctc acc tcc atg cag ctg aag acc ccc	1853
Leu Thr Leu Val Ser Arg Asp Leu Thr Ser Met Gln Leu Lys Thr Pro	
555 560 565	
agt ggc cag gtc ctc agc ttc tgc att ctg cag ctg ttt ccc ttc acc	1901
Ser Gly Gln Val Leu Ser Phe Cys Ile Leu Gln Leu Phe Pro Phe Thr	
570 575 580	
tcc gag agc aag cgg atg ggc gtc atc gtc agg gat gaa tcc acg gca	1949
Ser Glu Ser Lys Arg Met Gly Val Ile Val Arg Asp Glu Ser Thr Ala	
585 590 595	
gaa atc aca ttc tac atg aag ggc gct gac gtg gcc atg tct cct atc	1997
Glu Ile Thr Phe Tyr Met Lys Gly Ala Asp Val Ala Met Ser Pro Ile	
600 605 610	
gtg cag tat aat gac tgg ctg gaa gag gag tgc gga aac atg gct cgc	2045
Val Gln Tyr Asn Asp Trp Leu Glu Glu Glu Cys Gly Asn Met Ala Arg	
615 620 625 630	
gaa gga ctg cgg acc ctc gtg gtt gca aag aag cgc ttg aca gag gag	2093
Glu Gly Leu Arg Thr Leu Val Val Ala Lys Lys Ala Leu Thr Glu Glu	
635 640 645	

cag tac cag gac ttt gag agc cga tac act caa gcc aag ctg agc atg	2141
Gln Tyr Gln Asp Phe Glu Ser Arg Tyr Thr Gln Ala Lys Leu Ser Met	
650 655 660	
cac gac agg tcc ctc aag gtg gcc gcg gta gtc gag agc ctg gag agg	2189
His Asp Arg Ser Leu Lys Val Ala Val Val Glu Ser Leu Glu Arg	
665 670 675	
gag atg gaa ctg ctg tgc ctc acc ggc gtg gag gac cag ctg cag gca	2237
Glu Met Glu Leu Leu Cys Leu Thr Gly Val Glu Asp Gln Leu Gln Ala	
680 685 690	
gac gtg cgg ccc acg ctg gag atg ctg cgc aac gcc ggg atc aag ata	2285
Asp Val Arg Pro Thr Leu Glu Met Leu Arg Asn Ala Gly Ile Lys Ile	
695 700 705 710	
tgg atg cta aca ggc gat aaa ctc gag aca gct acc tgc att gcc aaa	2333
Trp Met Leu Thr Gly Asp Lys Leu Glu Thr Ala Thr Cys Ile Ala Lys	
715 720 725	
agt tca cat ctc gtg tct aga aca caa gat att cat att ttc aga cag	2381
Ser Ser His Leu Val Ser Arg Thr Gln Asp Ile His Ile Phe Arg Gln	
730 735 740	
gta acc agt cgg gga gag gca cat ttg gag ctg aat gca ttt cga agg	2429
Val Thr Ser Arg Gly Glu Ala His Leu Glu Leu Asn Ala Phe Arg Arg	
745 750 755	
aag cat gat tgt gca cta gtc ata tct ggg gac tct ctg gag gtt tgt	2477
Lys His Asp Cys Ala Leu Val Ile Ser Gly Asp Ser Leu Glu Val Cys	
760 765 770	
cta aag tac tac gag cat gaa ttt gtg gag ctg gcc tgc cag tgc cct	2525
Leu Lys Tyr Tyr Glu His Glu Phe Val Glu Leu Ala Cys Gln Cys Pro	
775 780 785 790	
gcc gtg gtt tgc tgc cgc tgc tca ccc acc cag aag gcc cgc att gtg	2573
Ala Val Val Cys Cys Arg Cys Ser Pro Thr Gln Lys Ala Arg Ile Val	
795 800 805	
aca ctg ctg cag cag cac aca ggg aga cgc acc tgc gcc atc ggt gat	2621
Thr Leu Leu Gln Gln His Thr Gly Arg Arg Thr Cys Ala Ile Gly Asp	
810 815 820	
gga gga aat gat gtc agc atg att cag gca gca gac tgt ggg att ggg	2669
Gly Gly Asn Asp Val Ser Met Ile Gln Ala Ala Asp Cys Gly Ile Gly	
825 830 835	
att gag gga aag gag ggt aaa cag gcc tcg ctg gcg gcc gac ttc tcc	2717
Ile Glu Gly Lys Glu Gly Lys Gln Ala Ser Leu Ala Ala Asp Phe Ser	
840 845 850	
atc acg cag ttc cgg cac ata ggc agg ctg ctc atg gtg cac ggg cgg	2765
Ile Thr Gln Phe Arg His Ile Gly Arg Leu Leu Met Val His Gly Arg	
855 860 865 870	
aac agc tac aag agg tcg gcg gca ctc ggc cag ttc gtc atg cac agg	2813
Asn Ser Tyr Lys Arg Ser Ala Ala Leu Gly Gln Phe Val Met His Arg	
875 880 885	

ggc ctt atc atc tcc acc atg cag gct gtg ttt tcc tca gtc ttc tac 2861
 Gly Leu Ile Ile Ser Thr Met Gln Ala Val Phe Ser Ser Val Phe Tyr
 890 895 900

ttc gca tcc gtc cct ttg tat cag ggc ttc ctc atg gtg ggg tat gcc 2909
 Phe Ala Ser Val Pro Leu Tyr Gln Gly Phe Leu Met Val Gly Tyr Ala
 905 910 915

acc ata tac acc atg ttc cca gtg ttc tcc tta gtg ctg gac cag gac 2957
 Thr Ile Tyr Thr Met Phe Pro Val Phe Ser Leu Val Leu Asp Gln Asp
 920 925 930

gtg aag cca gag atg gcg atg ctc tac ccg gag ctg tac aag gac ctc 3005
 Val Lys Pro Glu Met Ala Met Leu Tyr Pro Glu Leu Tyr Lys Asp Leu
 935 940 945 950

acc aag gga aga tcc ttg tcc ttc aaa acc ttc ctc atc tgg gtt tta 3053
 Thr Lys Gly Arg Ser Leu Ser Phe Lys Thr Phe Leu Ile Trp Val Leu
 955 960 965

ata agt att tac caa ggc ggc atc ctc atg tat ggg gcc ctg gtg ctc 3101
 Ile Ser Ile Tyr Gln Gly Gly Ile Leu Met Tyr Gly Ala Leu Val Leu
 970 975 980

ttc gag tct gag ttc gtc cac gtg gtg gcc atc tcc ttc acc gca ctg 3149
 Phe Glu Ser Glu Phe Val His Val Val Ala Ile Ser Phe Thr Ala Leu
 985 990 995

atc ctg acc gag ctg ctg atg gtg gcg ctg acc gtc cgc acg tgg cac 3197
 Ile Leu Thr Glu Leu Leu Met Val Ala Leu Thr Val Arg Thr Trp His
 1000 1005 1010

tgg ctg atg gtg gtg gcc gag ttc ctc agc tta ggc tgc tac gtg tcc 3245
 Trp Leu Met Val Val Ala Glu Phe Leu Ser Leu Gly Cys Tyr Val Ser
 1015 1020 1025 1030

tca ctc gct ttt ctc aat gaa tat ttt ggt ata ggc aga gtg tct ttt 3293
 Ser Leu Ala Phe Leu Asn Glu Tyr Phe Gly Ile Gly Arg Val Ser Phe
 1035 1040 1045

gga gct ttc tta gat gtt gcc ttt atc acc acc gtg acc ttc ctg tgg 3341
 Gly Ala Phe Leu Asp Val Ala Phe Ile Thr Thr Val Thr Phe Leu Trp
 1050 1055 1060

aaa gtg tgc gcg atc acc gtg gtc agc tgc ctc ccg ctg tat gtc ctc 3389
 Lys Val Ser Ala Ile Thr Val Val Ser Cys Leu Pro Leu Tyr Val Leu
 1065 1070 1075

aag tac ctg agg cgc aag tct tct cct ccc agc tac tgc aag ctg gcc 3437
 Lys Tyr Leu Arg Arg Lys Ser Ser Pro Pro Ser Tyr Cys Lys Leu Ala
 1080 1085 1090

tcc taa ggggctgtgc acccccagcg ggctggcccc agcaccttct gcccttccca 3493
 Ser *
 1095

gcaccttggtg cccctgccag tgaacgcagg gtttgccatt gctaccaagc aagcaccaca 3553
 agaaagggag ggtacgccag gcgagcccag ggcacagatg ctgagacagc ctctccttct 3613
 cagtgcaggg acgtcacccc tgccaggcaa gcccagggca cagatgccag gatggcttct 3673
 cctctcagtc gcgaggcttc acccctgccg ggcaagccca gggcatagat gctgagacag 3733
 cctctccctc tcagtgcagg gacgtcaccc ctgccaggca agcccagggc acagaggccg 3793

```

ggacggcctc tccctctcag tgtgaggctt caccatgct aggcaagccc agggcacaga 3853
tgccgggatg gccctccct cttagtgcgg gaaogtcacc cctgccaggc aagccoaggg 3913
cacagatget gcgatggcct cttcctctta agtgtggggc ctcacccctg cttttctttc 3973
tttttttgta ttgtcaaaat tgtatttcca tattgaagca gcttgagttt ctactgaaaa 4033
tgagcccgaa ttatttcaact attactgtaa aggggttcac ttactctggc attctgagaa 4093
ttagactgaa agtttaattt ctgcagtcc ctcattattca gattctttct ttgatgttat 4153
aacacaaagt catttctact caaatgtaat aaaattgagg ctccacggag aaaaaaaaaa 4213
aaaaaaaaaa aaaaaaaaaa 4231

```

<210> 26

<211> 1095

<212> PRT

<213> Homo sapiens

<400> 26

```

Met Pro Leu Met Met Ser Glu Glu Gly Phe Glu Asn Glu Glu Ser Asp
1      5      10      15
Tyr His Thr Leu Pro Arg Ala Arg Ile Met Gln Arg Lys Arg Gly Leu
20      25      30
Glu Trp Phe Val Cys Asp Gly Trp Lys Phe Leu Cys Thr Ser Cys Cys
35      40      45
Gly Trp Leu Ile Asn Ile Cys Arg Arg Lys Lys Glu Leu Lys Ala Arg
50      55      60
Thr Val Trp Leu Gly Cys Pro Glu Lys Cys Glu Glu Lys His Pro Arg
65      70      75      80
Asn Ser Ile Lys Asn Gln Lys Tyr Asn Val Phe Thr Phe Ile Pro Gly
85      90      95
Val Leu Tyr Glu Gln Phe Lys Phe Phe Leu Asn Leu Tyr Phe Leu Val
100      105      110
Ile Ser Cys Ser Gln Phe Val Pro Ala Leu Lys Ile Gly Tyr Leu Tyr
115      120      125
Thr Tyr Trp Ala Pro Leu Gly Phe Val Leu Ala Val Thr Met Thr Arg
130      135      140
Glu Ala Ile Asp Glu Phe Arg Arg Phe Gln Arg Asp Lys Glu Val Asn
145      150      155      160
Ser Gln Leu Tyr Ser Lys Leu Thr Val Arg Gly Lys Val Gln Val Lys
165      170      175
Ser Ser Asp Ile Gln Val Gly Asp Leu Ile Ile Val Glu Lys Asn Gln
180      185      190
Arg Ile Pro Ser Asp Met Val Phe Leu Arg Thr Ser Glu Lys Ala Gly
195      200      205
Ser Cys Phe Ile Arg Thr Asp Gln Leu Asp Gly Glu Thr Asp Trp Lys
210      215      220
Leu Lys Val Ala Val Ser Cys Thr Gln Gln Leu Pro Ala Leu Gly Asp
225      230      235      240
Leu Phe Ser Ile Ser Ala Tyr Val Tyr Ala Gln Lys Pro Gln Met Asp
245      250      255
Ile His Ser Phe Glu Gly Thr Phe Thr Arg Glu Asp Ser Asp Pro Pro
260      265      270
Ile His Glu Ser Leu Ser Ile Glu Asn Thr Leu Trp Ala Ser Thr Ile
275      280      285
Val Ala Ser Gly Thr Val Ile Gly Val Val Ile Tyr Thr Gly Lys Glu
290      295      300
Thr Arg Ser Val Met Asn Thr Ser Asn Pro Lys Asn Lys Val Gly Leu
305      310      315      320
Leu Asp Leu Glu Leu Asn Arg Leu Thr Lys Ala Leu Phe Leu Ala Leu
325      330      335
Val Ala Leu Ser Ile Val Met Val Thr Leu Gln Gly Phe Val Gly Pro
340      345      350
Trp Tyr Arg Asn Leu Phe Arg Phe Leu Leu Leu Phe Ser Tyr Ile Ile
355      360      365

```

Pro	Ile	Ser	Leu	Arg	Val	Asn	Leu	Asp	Met	Gly	Lys	Ala	Val	Tyr	Gly	370	375	380
Trp	Met	Met	Met	Lys	Asp	Glu	Asn	Ile	Pro	Gly	Thr	Val	Val	Arg	Thr	385	390	395
Ser	Thr	Ile	Pro	Glu	Glu	Leu	Gly	Arg	Leu	Val	Tyr	Leu	Leu	Thr	Asp	405	410	415
Lys	Thr	Gly	Thr	Leu	Thr	Gln	Asn	Glu	Met	Ile	Phe	Lys	Arg	Leu	His	420	425	430
Leu	Gly	Thr	Val	Ser	Tyr	Gly	Ala	Asp	Thr	Met	Asp	Glu	Ile	Gln	Ser	435	440	445
His	Val	Arg	Asp	Ser	Tyr	Ser	Gln	Met	Gln	Ser	Gln	Ala	Gly	Gly	Asn	450	455	460
Asn	Thr	Gly	Ser	Thr	Pro	Leu	Arg	Lys	Ala	Gln	Ser	Ser	Ala	Pro	Lys	465	470	475
Val	Arg	Lys	Ser	Val	Ser	Ser	Arg	Ile	His	Glu	Ala	Val	Lys	Ala	Ile	485	490	495
Val	Leu	Cys	His	Asn	Val	Thr	Pro	Val	Tyr	Glu	Ser	Arg	Ala	Gly	Val	500	505	510
Thr	Glu	Glu	Thr	Glu	Phe	Ala	Glu	Ala	Asp	Gln	Asp	Phe	Ser	Asp	Glu	515	520	525
Asn	Arg	Thr	Tyr	Gln	Ala	Ser	Ser	Pro	Asp	Glu	Val	Ala	Leu	Val	Gln	530	535	540
Trp	Thr	Glu	Ser	Val	Gly	Leu	Thr	Leu	Val	Ser	Arg	Asp	Leu	Thr	Ser	545	550	555
Met	Gln	Leu	Lys	Thr	Pro	Ser	Gly	Gln	Val	Leu	Ser	Phe	Cys	Ile	Leu	565	570	575
Gln	Leu	Phe	Pro	Phe	Thr	Ser	Glu	Ser	Lys	Arg	Met	Gly	Val	Ile	Val	580	585	590
Arg	Asp	Glu	Ser	Thr	Ala	Glu	Ile	Thr	Phe	Tyr	Met	Lys	Gly	Ala	Asp	595	600	605
Val	Ala	Met	Ser	Pro	Ile	Val	Gln	Tyr	Asn	Asp	Trp	Leu	Glu	Glu	Glu	610	615	620
Cys	Gly	Asn	Met	Ala	Arg	Glu	Gly	Leu	Arg	Thr	Leu	Val	Val	Ala	Lys	625	630	635
Lys	Ala	Leu	Thr	Glu	Glu	Gln	Tyr	Gln	Asp	Phe	Glu	Ser	Arg	Tyr	Thr	645	650	655
Gln	Ala	Lys	Leu	Ser	Met	His	Asp	Arg	Ser	Leu	Lys	Val	Ala	Ala	Val	660	665	670
Val	Glu	Ser	Leu	Glu	Arg	Glu	Met	Glu	Leu	Leu	Cys	Leu	Thr	Gly	Val	675	680	685
Glu	Asp	Gln	Leu	Gln	Ala	Asp	Val	Arg	Pro	Thr	Leu	Glu	Met	Leu	Arg	690	695	700
Asn	Ala	Gly	Ile	Lys	Ile	Trp	Met	Leu	Thr	Gly	Asp	Lys	Leu	Glu	Thr	705	710	715
Ala	Thr	Cys	Ile	Ala	Lys	Ser	Ser	His	Leu	Val	Ser	Arg	Thr	Gln	Asp	725	730	735
Ile	His	Ile	Phe	Arg	Gln	Val	Thr	Ser	Arg	Gly	Glu	Ala	His	Leu	Glu	740	745	750
Leu	Asn	Ala	Phe	Arg	Arg	Lys	His	Asp	Cys	Ala	Leu	Val	Ile	Ser	Gly	755	760	765
Asp	Ser	Leu	Glu	Val	Cys	Leu	Lys	Tyr	Tyr	Glu	His	Glu	Phe	Val	Glu	770	775	780
Leu	Ala	Cys	Gln	Cys	Pro	Ala	Val	Val	Cys	Cys	Arg	Cys	Ser	Pro	Thr	785	790	795
Gln	Lys	Ala	Arg	Ile	Val	Thr	Leu	Leu	Gln	Gln	His	Thr	Gly	Arg	Arg	805	810	815
Thr	Cys	Ala	Ile	Gly	Asp	Gly	Gly	Asn	Asp	Val	Ser	Met	Ile	Gln	Ala	820	825	830
Ala	Asp	Cys	Gly	Ile	Gly	Ile	Glu	Gly	Lys	Glu	Gly	Lys	Gln	Ala	Ser	835	840	845

Leu Ala Ala Asp Phe Ser Ile Thr Gln Phe Arg His Ile Gly Arg Leu
 850 855 860
 Leu Met Val His Gly Arg Asn Ser Tyr Lys Arg Ser Ala Ala Leu Gly
 865 870 875 880
 Gln Phe Val Met His Arg Gly Leu Ile Ile Ser Thr Met Gln Ala Val
 885 890 895
 Phe Ser Ser Val Phe Tyr Phe Ala Ser Val Pro Leu Tyr Gln Gly Phe
 900 905 910
 Leu Met Val Gly Tyr Ala Thr Ile Tyr Thr Met Phe Pro Val Phe Ser
 915 920 925
 Leu Val Leu Asp Gln Asp Val Lys Pro Glu Met Ala Met Leu Tyr Pro
 930 935 940
 Glu Leu Tyr Lys Asp Leu Thr Lys Gly Arg Ser Leu Ser Phe Lys Thr
 945 950 955 960
 Phe Leu Ile Trp Val Leu Ile Ser Ile Tyr Gln Gly Gly Ile Leu Met
 965 970 975
 Tyr Gly Ala Leu Val Leu Phe Glu Ser Glu Phe Val His Val Val Ala
 980 985 990
 Ile Ser Phe Thr Ala Leu Ile Leu Thr Glu Leu Leu Met Val Ala Leu
 995 1000 1005
 Thr Val Arg Thr Trp His Trp Leu Met Val Val Ala Glu Phe Leu Ser
 1010 1015 1020
 Leu Gly Cys Tyr Val Ser Ser Leu Ala Phe Leu Asn Glu Tyr Phe Gly
 1025 1030 1035 1040
 Ile Gly Arg Val Ser Phe Gly Ala Phe Leu Asp Val Ala Phe Ile Thr
 1045 1050 1055
 Thr Val Thr Phe Leu Trp Lys Val Ser Ala Ile Thr Val Val Ser Cys
 1060 1065 1070
 Leu Pro Leu Tyr Val Leu Lys Tyr Leu Arg Arg Lys Ser Ser Pro Pro
 1075 1080 1085
 Ser Tyr Cys Lys Leu Ala Ser
 1090 1095

<210> 27

<211> 3255

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)...(3255)

<400> 27

atg cca cta atg atg tct gaa gaa ggc ttt gag aat gag gaa agt gat 48
 Met Pro Leu Met Met Ser Glu Glu Gly Phe Glu Asn Glu Glu Ser Asp
 1 5 10 15
 tac cac acc tta cca cga gcc agg ata atg caa agg aaa aga gga ctg 96
 Tyr His Thr Leu Pro Arg Ala Arg Ile Met Gln Arg Lys Arg Gly Leu
 20 25 30
 gag tgg ttt gtc tgt gat ggc tgg aag ttc ctc tgt acc agt tgc tgt 144
 Glu Trp Phe Val Cys Asp Gly Trp Lys Phe Leu Cys Thr Ser Cys Cys
 35 40 45
 ggt tgg ctg ata aat att tgt cga aga aag aaa gag ctg aaa gct cgc 192
 Gly Trp Leu Ile Asn Ile Cys Arg Arg Lys Lys Glu Leu Lys Ala Arg
 50 55 60

aca gta tgg ctt gga tgt cct gaa aag tgt gaa gaa aaa cat ccc agg	240
Thr Val Trp Leu Gly Cys Pro Glu Lys Cys Glu Glu Lys His Pro Arg	
65 70 75 80	
aat tct ata aaa aat caa aaa tac aat gtg ttt acc ttt ata cct ggg	288
Asn Ser Ile Lys Asn Gln Lys Tyr Asn Val Phe Thr Phe Ile Pro Gly	
85 90 95	
gtt ttg tat gaa caa ttc aag ttt ttc ttg aat ctc tat ttt cta gtg	336
Val Leu Tyr Glu Gln Phe Lys Phe Phe Leu Asn Leu Tyr Phe Leu Val	
100 105 110	
ata tcc tgc tca cag ttt gta cca gca ttg aaa ata ggc tat ctc tac	384
Ile Ser Cys Ser Gln Phe Val Pro Ala Leu Lys Ile Gly Tyr Leu Tyr	
115 120 125	
acc tac tgg gct cct ctg gga ttt gtc ttg gct gtt act atg aca cgg	432
Thr Tyr Trp Ala Pro Leu Gly Phe Val Leu Ala Val Thr Met Thr Arg	
130 135 140	
gaa gca att gat gaa ttt cgg cgt ttt cag cgt gac aag gaa gtg aat	480
Glu Ala Ile Asp Glu Phe Arg Arg Phe Gln Arg Asp Lys Glu Val Asn	
145 150 155 160	
tca caa cta tat agc aag ctt aca gta aga ggt aaa gtg caa gtt aag	528
Ser Gln Leu Tyr Ser Lys Leu Thr Val Arg Gly Lys Val Gln Val Lys	
165 170 175	
agt tca gac ata caa gtt gga gac ctc atc ata gtg gaa aag aat caa	576
Ser Ser Asp Ile Gln Val Gly Asp Leu Ile Ile Val Glu Lys Asn Gln	
180 185 190	
aga att cca tcg gac atg gtg ttt ctt agg act tca gaa aaa gca ggt	624
Arg Ile Pro Ser Asp Met Val Phe Leu Arg Thr Ser Glu Lys Ala Gly	
195 200 205	
tcg tgt ttt att cga act gat caa cta gat ggt gaa act gac tgg aag	672
Ser Cys Phe Ile Arg Thr Asp Gln Leu Asp Gly Glu Thr Asp Trp Lys	
210 215 220	
ctg aag gtg gca gtg agc tgc acg caa cag ctg ccg gct ctg ggg gac	720
Leu Lys Val Ala Val Ser Cys Thr Gln Gln Leu Pro Ala Leu Gly Asp	
225 230 235 240	
ctt ttt tct atc agt gct tat gtt tat gct cag aaa cca caa atg gac	768
Leu Phe Ser Ile Ser Ala Tyr Val Tyr Ala Gln Lys Pro Gln Met Asp	
245 250 255	
att cac agt ttc gaa ggc aca ttt acc agg gaa gac agt gac ccg ccc	816
Ile His Ser Phe Glu Gly Thr Phe Thr Arg Glu Asp Ser Asp Pro Pro	
260 265 270	
att cat gaa agt ctc agc ata gaa aat aca ttg tgg gca agc acc att	864
Ile His Glu Ser Leu Ser Ile Glu Asn Thr Leu Trp Ala Ser Thr Ile	
275 280 285	
gtt gca tca ggt act gta ata ggt gtt gtc att tat acc gga aaa gag	912
Val Ala Ser Gly Thr Val Ile Gly Val Val Ile Tyr Thr Gly Lys Glu	
290 295 300	

act cga agt gta atg aac aca tcc aat cca aaa aat aag gtt ggt ttg	960
Thr Arg Ser Val Met Asn Thr Ser Asn Pro Lys Asn Lys Val Gly Leu	
305 310 315 320	
ttg gac ctt gaa ctc aat cgg ctg acg aaa gcg cta ttt ttg gct tta	1008
Leu Asp Leu Glu Leu Asn Arg Leu Thr Lys Ala Leu Phe Leu Ala Leu	
325 330 335	
gtt gct ctt tcc att gtt atg gta acc tta caa gga ttt gtg ggt cca	1056
Val Ala Leu Ser Ile Val Met Val Thr Leu Gln Gly Phe Val Gly Pro	
340 345 350	
tgg tac cgc aat ctt ttt cgg ttc ctt ctc ctc ttt tct tac atc att	1104
Trp Tyr Arg Asn Leu Phe Arg Phe Leu Leu Leu Phe Ser Tyr Ile Ile	
355 360 365	
ccc ata agt ttg cgt gtg aac ttg gac atg ggc aaa gcg gtg tat gga	1152
Pro Ile Ser Leu Arg Val Asn Leu Asp Met Gly Lys Ala Val Tyr Gly	
370 375 380	
tgg atg atg atg aaa gat gag aac atc cct ggc acg gtc gtt cgg acc	1200
Trp Met Met Met Lys Asp Glu Asn Ile Pro Gly Thr Val Val Arg Thr	
385 390 395 400	
agc act atc cca gag gaa ctt ggg cgc ctg gtg tat tta ttg aca gac	1248
Ser Thr Ile Pro Glu Glu Leu Gly Arg Leu Val Tyr Leu Leu Thr Asp	
405 410 415	
aaa aca gga acc ctc acc cag aat gaa atg ata ttt aag cgg ctg cac	1296
Lys Thr Gly Thr Leu Thr Gln Asn Glu Met Ile Phe Lys Arg Leu His	
420 425 430	
ctg ggc acc gtg tcc tat ggc gcc gac acg atg gat gag atc cag agc	1344
Leu Gly Thr Val Ser Tyr Gly Ala Asp Thr Met Asp Glu Ile Gln Ser	
435 440 445	
cat gtc agg gac tcc tac tca cag atg cag tct caa gct ggt gga aac	1392
His Val Arg Asp Ser Tyr Ser Gln Met Gln Ser Gln Ala Gly Gly Asn	
450 455 460	
aat act ggt tca act cca cta aga aaa gcc caa tct tca gct ccc aaa	1440
Asn Thr Gly Ser Thr Pro Leu Arg Lys Ala Gln Ser Ser Ala Pro Lys	
465 470 475 480	
gtt agg aaa agt gtc agt agt cga atc cat gaa gcc gtg aaa gcc atc	1488
Val Arg Lys Ser Val Ser Ser Arg Ile His Glu Ala Val Lys Ala Ile	
485 490 495	
gtg ctg tgt cac aac gtg acc ccc gtg tat gag tct cgg gcc ggc gtt	1536
Val Leu Cys His Asn Val Thr Pro Val Tyr Glu Ser Arg Ala Gly Val	
500 505 510	
act gag gag act gag ttc gca gag gct gac caa gac ttc agt gat gag	1584
Thr Glu Glu Thr Glu Phe Ala Glu Ala Asp Gln Asp Phe Ser Asp Glu	
515 520 525	
aat cgc acc tac cag gct tcc agc ccg gat gag gtc gct ctg gtg cag	1632
Asn Arg Thr Tyr Gln Ala Ser Ser Pro Asp Glu Val Ala Leu Val Gln	
530 535 540	

tgg aca gag agt gtg ggc ctc acg ctg gtc agc agg gac ctc acc tcc	1680
Trp Thr Glu Ser Val Gly Leu Thr Leu Val Ser Arg Asp Leu Thr Ser	
545 550 555 560	
atg cag ctg aag acc ccc agt ggc cag gtc ctc agc ttc tgc att ctg	1728
Met Gln Leu Lys Thr Pro Ser Gly Gln Val Leu Ser Phe Cys Ile Leu	
565 570 575	
cag ctg ttt ccc ttc acc tcc gag agc aag cgg atg ggc gtc atc gtc	1776
Gln Leu Phe Pro Phe Thr Ser Glu Ser Lys Arg Met Gly Val Ile Val	
580 585 590	
agg gat gaa tcc acg gca gaa atc aca ttc tac atg aag ggc gct gac	1824
Arg Asp Glu Ser Thr Ala Glu Ile Thr Phe Tyr Met Lys Gly Ala Asp	
595 600 605	
gtg gcc atg tct cct atc gtg cag tat aat gac tgg ctg gaa gag gag	1872
Val Ala Met Ser Pro Ile Val Gln Tyr Asn Asp Trp Leu Glu Glu Glu	
610 615 620	
tgc gga aac atg gct cgc gaa gga ctg cgg acc ctc gtg gtt gca aag	1920
Cys Gly Asn Met Ala Arg Glu Gly Leu Arg Thr Leu Val Val Ala Lys	
625 630 635 640	
aag gcg ttg aca gag gag cag tac cag gac ttt gag agc cga tac act	1968
Lys Ala Leu Thr Glu Glu Gln Tyr Gln Asp Phe Glu Ser Arg Tyr Thr	
645 650 655	
caa gcc aag ctg agc atg cac gac agg tcc ctc aag gtg gcc gcg gta	2016
Gln Ala Lys Leu Ser Met His Asp Arg Ser Leu Lys Val Ala Ala Val	
660 665 670	
gtc gag agc ctg gag agg gag atg gaa ctg ctg tgc ctc acc ggc gtg	2064
Val Glu Ser Leu Glu Arg Glu Met Glu Leu Leu Cys Leu Thr Gly Val	
675 680 685	
gag gac cag ctg cag gca gac gtg cgg ccc acg ctg gag atg ctg cgc	2112
Glu Asp Gln Leu Gln Ala Asp Val Arg Pro Thr Leu Glu Met Leu Arg	
690 695 700	
aac gcc ggg atc aag ata tgg atg cta aca ggc gat aaa ctc gag aca	2160
Asn Ala Gly Ile Lys Ile Trp Met Leu Thr Gly Asp Lys Leu Glu Thr	
705 710 715 720	
gct acc tgc att gcc aaa agt tca cat ctc gtg tct aga aca caa gat	2208
Ala Thr Cys Ile Ala Lys Ser Ser His Leu Val Ser Arg Thr Gln Asp	
725 730 735	
att cat att ttc aga cag gta acc agt cgg gga gag gca cat ttg gag	2256
Ile His Ile Phe Arg Gln Val Thr Ser Arg Gly Glu Ala His Leu Glu	
740 745 750	
ctg aat gca ttt cga agg aag cat gat tgt gca cta gtc ata tct ggg	2304
Leu Asn Ala Phe Arg Arg Lys His Asp Cys Ala Leu Val Ile Ser Gly	
755 760 765	
gac tct ctg gag gtt tgt cta aag tac tac gag cat gaa ttt gtg gag	2352
Asp Ser Leu Glu Val Cys Leu Lys Tyr Tyr Glu His Glu Phe Val Glu	
770 775 780	

ctg gcc tgc cag tgc cct gcc gtg gtt tgc tgc cgc tgc tca ccc acc	2400
Leu Ala Cys Gln Cys Pro Ala Val Val Cys Cys Arg Cys Ser Pro Thr	
785 790 795 800	
cag aag gcc cgc att gtg aca ctg ctg cag cag cac aca ggg aga cgc	2448
Gln Lys Ala Arg Ile Val Thr Leu Leu Gln Gln His Thr Gly Arg Arg	
805 810 815	
acc tgc gcc atc ggt gat gga gga aat gat gtc agc atg att cag gca	2496
Thr Cys Ala Ile Gly Asp Gly Gly Asn Asp Val Ser Met Ile Gln Ala	
820 825 830	
gca gac tgt ggg att ggg att gag gga aag gag ggt aaa cag gcc tcg	2544
Ala Asp Cys Gly Ile Gly Ile Glu Gly Lys Glu Gly Lys Gln Ala Ser	
835 840 845	
ctg gcg gcc gac ttc tcc atc acg cag ttc cgg cac ata ggc agg ctg	2592
Leu Ala Ala Asp Phe Ser Ile Thr Gln Phe Arg His Ile Gly Arg Leu	
850 855 860	
ctc atg gtg cac ggg cgg aac agc tac aag agg tcg gcg gca ctc ggc	2640
Leu Met Val His Gly Arg Asn Ser Tyr Lys Arg Ser Ala Ala Leu Gly	
865 870 875 880	
cag ttc gtc atg cac agg ggc ctt atc atc tcc acc atg cag gct gtg	2688
Gln Phe Val Met His Arg Gly Leu Ile Ile Ser Thr Met Gln Ala Val	
885 890 895	
ttt tcc tca gtc ttc tac ttc gca tcc gtc cct ttg tat cag ggc ttc	2736
Phe Ser Ser Val Phe Tyr Phe Ala Ser Val Pro Leu Tyr Gln Gly Phe	
900 905 910	
ctc atg gtg ggg tat gcc acc ata tac acc atg ttc cca gtg ttc tcc	2784
Leu Met Val Gly Tyr Ala Thr Ile Tyr Thr Met Phe Pro Val Phe Ser	
915 920 925	
tta gtg ctg gac cag gac gtg aag cca gag atg gcg atg ctc tac ccg	2832
Leu Val Leu Asp Gln Asp Val Lys Pro Glu Met Ala Met Leu Tyr Pro	
930 935 940	
gag ctg tac aag gac ctc acc aag gga aga tcc ttg tcc ttc aaa acc	2880
Glu Leu Tyr Lys Asp Leu Thr Lys Gly Arg Ser Leu Ser Phe Lys Thr	
945 950 955 960	
ttc ctc atc tgg gtt tta ata agt att tac caa ggc ggc atc ctc atg	2928
Phe Leu Ile Trp Val Leu Ile Ser Ile Tyr Gln Gly Gly Ile Leu Met	
965 970 975	
tat ggg gcc ctg gtg ctc ttc gag tct gag ttc gtc cac gtg gtg gcc	2976
Tyr Gly Ala Leu Val Leu Phe Glu Ser Glu Phe Val His Val Val Ala	
980 985 990	
atc tcc ttc acc gca ctg atc ctg acc gag ctg ctg atg gtg gcg ctg	3024
Ile Ser Phe Thr Ala Leu Ile Leu Thr Glu Leu Leu Met Val Ala Leu	
995 1000 1005	
acc gtc cgc acg tgg cac tgg ctg atg gtg gtg gcc gag ttc ctc agc	3072
Thr Val Arg Thr Trp His Trp Leu Met Val Val Ala Glu Phe Leu Ser	
1010 1015 1020	

tta ggc tgc tac gtg tcc tca ctc gct ttt ctc aat gaa tat ttt gat 3120
 Leu Gly Cys Tyr Val Ser Ser Leu Ala Phe Leu Asn Glu Tyr Phe Asp
 1025 1030 1035 1040

gtt gcc ttt atc acc acc gtg acc ttc ctg tgg aaa gtg tgc gcg atc 3168
 Val Ala Phe Ile Thr Thr Val Thr Phe Leu Trp Lys Val Ser Ala Ile
 1045 1050 1055

acc gtg gtc agc tgc ctc ccg ctg tat gtc ctc aag tac ctg agg cgc 3216
 Thr Val Val Ser Cys Leu Pro Leu Tyr Val Leu Lys Tyr Leu Arg Arg
 1060 1065 1070

aag tct tct cct ccc agc tac tgc aag ctg gcc tcc taa 3255
 Lys Ser Ser Pro Ser Tyr Cys Lys Leu Ala Ser *
 1075 1080

<210> 28

<211> 464

<212> PRT

<213> Escherichia coli

<400> 28

Met Pro Asp Ala Lys Lys Gln Gly Arg Ser Asn Lys Ala Met Thr Phe
 1 5 10 15
 Phe Val Cys Phe Leu Ala Ala Leu Ala Gly Leu Leu Phe Gly Leu Asp
 20 25 30
 Ile Gly Val Ile Ala Gly Ala Leu Pro Phe Ile Ala Asp Glu Phe Gln
 35 40 45
 Ile Thr Ser His Thr Gln Glu Trp Val Val Ser Ser Met Met Phe Gly
 50 55 60
 Ala Ala Val Gly Ala Val Gly Ser Gly Trp Leu Ser Phe Lys Leu Gly
 65 70 75 80
 Arg Lys Lys Ser Leu Met Ile Gly Ala Ile Leu Phe Val Ala Gly Ser
 85 90 95
 Leu Phe Ser Ala Ala Ala Pro Asn Val Glu Val Leu Ile Leu Ser Arg
 100 105 110
 Val Leu Leu Gly Leu Ala Val Gly Val Ala Ser Tyr Thr Ala Pro Leu
 115 120 125
 Tyr Leu Ser Glu Ile Ala Pro Glu Lys Ile Arg Gly Ser Met Ile Ser
 130 135 140
 Met Tyr Gln Leu Met Ile Thr Ile Gly Ile Leu Gly Ala Tyr Leu Ser
 145 150 155 160
 Asp Thr Ala Phe Ser Tyr Thr Gly Ala Trp Arg Trp Met Leu Gly Val
 165 170 175
 Ile Ile Ile Pro Ala Ile Leu Leu Leu Ile Gly Val Phe Phe Leu Pro
 180 185 190
 Asp Ser Pro Arg Trp Phe Ala Ala Lys Arg Arg Phe Val Asp Ala Glu
 195 200 205
 Arg Val Leu Leu Arg Leu Arg Asp Thr Ser Ala Glu Ala Lys Arg Glu
 210 215 220
 Leu Asp Glu Ile Arg Glu Ser Leu Gln Val Lys Gln Ser Gly Trp Ala
 225 230 235 240
 Leu Phe Lys Glu Asn Ser Asn Phe Arg Arg Ala Val Phe Leu Gly Val
 245 250 255
 Leu Leu Gln Val Met Gln Gln Phe Thr Gly Met Asn Val Ile Met Tyr
 260 265 270
 Tyr Ala Pro Lys Ile Phe Glu Leu Ala Gly Tyr Thr Asn Thr Thr Glu
 275 280 285
 Gln Met Trp Gly Thr Val Ile Val Gly Leu Thr Asn Val Leu Ala Thr
 290 295 300

```

Phe Ile Ala Ile Gly Leu Val Asp Arg Trp Gly Arg Lys Pro Thr Leu
305          310          315          320
Thr Leu Gly Phe Leu Val Met Ala Ala Gly Met Gly Val Leu Gly Thr
          325          330          335
Met Met His Ile Gly Ile His Ser Pro Ser Ala Gln Tyr Phe Ala Ile
          340          345          350
Ala Met Leu Leu Met Phe Ile Val Gly Phe Ala Met Ser Ala Gly Pro
          355          360          365
Leu Ile Trp Val Leu Cys Ser Glu Ile Gln Pro Leu Lys Gly Arg Asp
          370          375          380
Phe Gly Ile Thr Cys Ser Thr Ala Thr Asn Trp Ile Ala Asn Met Ile
385          390          395          400
Val Gly Ala Thr Phe Leu Thr Met Leu Asn Thr Leu Gly Asn Ala Asn
          405          410          415
Thr Phe Trp Val Tyr Ala Ala Leu Asn Val Leu Phe Ile Leu Leu Thr
          420          425          430
Leu Trp Leu Val Pro Glu Thr Lys His Val Ser Leu Glu His Ile Glu
          435          440          445
Arg Asn Leu Met Lys Gly Arg Lys Leu Arg Glu Ile Gly Ala His Asp
          450          455          460

```

<210> 29

<211> 472

<212> PRT

<213> Escherichia coli

<400> 29

```

Met Val Thr Ile Asn Thr Glu Ser Ala Leu Thr Pro Arg Ser Leu Arg
1          5          10          15
Asp Thr Arg Arg Met Asn Met Phe Val Ser Val Ala Ala Ala Val Ala
          20          25          30
Gly Leu Leu Phe Gly Leu Asp Ile Gly Val Ile Ala Gly Ala Leu Pro
          35          40          45
Phe Ile Thr Asp His Phe Val Leu Thr Ser Arg Leu Gln Glu Trp Val
          50          55          60
Val Ser Ser Met Met Leu Gly Ala Ala Ile Gly Ala Leu Phe Asn Gly
65          70          75          80
Trp Leu Ser Phe Arg Leu Gly Arg Lys Tyr Ser Leu Met Ala Gly Ala
          85          90          95
Ile Leu Phe Val Leu Gly Ser Ile Gly Ser Ala Phe Ala Thr Ser Val
          100          105          110
Glu Met Leu Ile Ala Ala Arg Val Val Leu Gly Ile Ala Val Gly Ile
          115          120          125
Ala Ser Tyr Thr Ala Pro Leu Tyr Leu Ser Glu Met Ala Ser Glu Asn
          130          135          140
Val Arg Gly Lys Met Ile Ser Met Tyr Gln Leu Met Val Thr Leu Gly
145          150          155          160
Ile Val Leu Ala Phe Leu Ser Asp Thr Ala Phe Ser Tyr Ser Gly Asn
          165          170          175
Trp Arg Ala Met Leu Gly Val Leu Ala Leu Pro Ala Val Leu Leu Ile
          180          185          190
Ile Leu Val Val Phe Leu Pro Asn Ser Pro Arg Trp Leu Ala Glu Lys
          195          200          205
Gly Arg His Ile Glu Ala Glu Glu Val Leu Arg Met Leu Arg Asp Thr
          210          215          220
Ser Glu Lys Ala Arg Glu Glu Leu Asn Glu Ile Arg Glu Ser Leu Lys
225          230          235          240
Leu Lys Gln Gly Gly Trp Ala Leu Phe Lys Ile Asn Arg Asn Val Arg
          245          250          255

```

```

Arg Ala Val Phe Leu Gly Met Leu Leu Gln Ala Met Gln Gln Phe Thr
      260      265      270
Gly Met Asn Ile Ile Met Tyr Tyr Ala Pro Arg Ile Phe Lys Met Ala
      275      280      285
Gly Phe Thr Thr Thr Glu Gln Met Ile Ala Thr Leu Val Val Gly
      290      295      300
Leu Thr Phe Met Phe Ala Thr Phe Ile Ala Val Phe Thr Val Asp Lys
      305      310      315      320
Ala Gly Arg Lys Pro Ala Leu Lys Ile Gly Phe Ser Val Met Ala Leu
      325      330      335
Gly Thr Leu Val Leu Gly Tyr Cys Leu Met Gln Phe Asp Asn Gly Thr
      340      345      350
Ala Ser Ser Gly Leu Ser Trp Leu Ser Val Gly Met Thr Met Met Cys
      355      360      365
Ile Ala Gly Tyr Ala Met Ser Ala Ala Pro Val Val Trp Ile Leu Cys
      370      375      380
Ser Glu Ile Gln Pro Leu Lys Cys Arg Asp Phe Gly Ile Thr Cys Ser
      385      390      395      400
Thr Thr Thr Asn Trp Val Ser Asn Met Ile Ile Gly Ala Thr Phe Leu
      405      410      415
Thr Leu Leu Asp Ser Ile Gly Ala Ala Gly Thr Phe Trp Leu Tyr Thr
      420      425      430
Ala Leu Asn Ile Ala Phe Val Gly Ile Thr Phe Trp Leu Ile Pro Glu
      435      440      445
Thr Lys Asn Val Thr Leu Glu His Ile Glu Arg Lys Leu Met Ala Gly
      450      455      460
Glu Lys Leu Arg Asn Ile Gly Val
      465      470

```

<210> 30

<211> 526

<212> PRT

<213> Homo sapiens

<400> 30

```

Met Thr Ser Asp His Glu His Met Thr Ala Val Cys Ala Ser His Val
  1      5      10      15
Gln Thr His Gly Ser Gln Leu Gln Ile Gln Lys Leu Ser Pro Cys Phe
      20      25      30
Arg Pro Pro Thr Pro Ala Phe Arg Ile Ser Ser Ser Ile Ile Leu Leu
      35      40      45
Gly Ala Gly Leu Ala Gly Pro Ser Thr Gly Asp Arg Trp Phe Gly Val
      50      55      60
Ser Val Val Gly Thr Gly Leu Phe Leu Pro Pro Leu Gln Leu Leu Leu
      65      70      75      80
Pro Pro Arg Leu Leu Phe Thr His Ala Ile Leu Glu Arg Leu His Leu
      85      90      95
Trp Leu Ala Leu Pro Pro Val Leu Val Leu Gly His Ala Leu Leu His
      100      105      110
Cys Lys Val Gly Gly Ser Thr Ala Arg Ala Gly Asp Gln Leu Val Gln
      115      120      125
Arg Val Leu Leu Leu Ile Val Phe Leu His Arg Trp Val Gln Val Trp
      130      135      140
Pro Glx Gly Thr Glu Val Asp Ile Leu Gly Met Gly Ser Arg Thr Gly
      145      150      155      160
Gly Arg Arg Gly Pro Glu Leu Arg Pro Gly Phe Arg Ile Ser Ile Leu
      165      170      175
Ser Ala Tyr Ile Ser Asn Tyr Ala Phe Ala Asn Val Phe His Gly Trp
      180      185      190

```

Lys Tyr Met Phe Gly Leu Val Ile Pro Leu Gly Val Leu Gln Ala Ile
 195 "200 205
 Ala Met Tyr Phe Leu Pro Pro Ser Pro Arg Phe Leu Val Met Lys Gly
 210 215 220
 Gln Glu Gly Ala Ala Ser Lys Val Leu Gly Arg Leu Arg Ala Leu Ser
 225 230 235 240
 Asp Thr Thr Glu Glu Leu Thr Val Ile Lys Ser Ser Leu Lys Asp Glu
 245 250 255
 Tyr Gln Tyr Ser Phe Trp Asp Leu Phe Arg Ser Lys Asp Asn Met Arg
 260 265 270
 Thr Arg Ile Met Ile Gly Leu Thr Leu Val Phe Phe Val Gln Ile Thr
 275 280 285
 Gly Gln Pro Asn Ile Leu Phe Tyr Ala Ser Thr Val Leu Lys Ser Val
 290 295 300
 Gly Phe Gln Ser Asn Glu Ala Ala Ser Leu Ala Ser Thr Gly Val Gly
 305 310 315 320
 Val Val Lys Val Ile Ser Thr Ile Pro Ala Thr Leu Leu Val Asp His
 325 330 335
 Val Gly Ser Lys Thr Phe Leu Cys Ile Gly Leu Leu Asn Ala Gly Leu
 340 345 350
 Ser His Thr Glu Tyr Gln Ile Val Thr Asp Pro Gly Asp Val Pro Ala
 355 360 365
 Phe Leu Lys Trp Leu Ser Leu Ala Ser Leu Leu Val Tyr Val Ala Ala
 370 375 380
 Phe Ser Ile Gly Leu Gly Pro Met Pro Trp Leu Val Leu Ser Glu Ile
 385 390 395 400
 Phe Pro Gly Gly Ile Arg Gly Arg Ala Met Ala Leu Thr Ser Ser Met
 405 410 415
 Asn Trp Gly Ile Asn Leu Leu Ile Ser Leu Thr Phe Leu Thr Val Asn
 420 425 430
 Leu Ile Gly Leu Pro Trp Val Cys Phe Ile Tyr Thr Ile Met Ser Leu
 435 440 445
 Ala Ser Leu Leu Phe Val Val Met Phe Ile Pro Glu Thr Lys Gly Cys
 450 455 460
 Ser Leu Glu Gln Ile Ser Met Glu Leu Ala Lys Val Asn Tyr Val Lys
 465 470 475 480
 Asn Asn Ile Cys Phe Met Ser His His Gln Glu Glu Leu Val Pro Lys
 485 490 495
 Gln Pro Gln Lys Arg Lys Pro Gln Glu Gln Leu Leu Glu Cys Asn Lys
 500 505 510
 Leu Cys Gly Arg Gly Gln Ser Arg Gln Leu Ser Pro Glu Thr
 515 520 525

<210> 31

<211> 25002

<212> DNA

<213> Caenorhabditis elegans

<400> 31

accctgcaat aggatatcat tacgatgtaa gtattcaata tttatttcaa atttgatatt 60
 tgttgacaaa ttttcaaata atgaaatgtg aatgcttcat gaattcttaa gctttcgggg 120
 cacagaaagt gagggcgatc ttcatittaa ctgcaaaatt tcaggtacca tcatgtaatg 180
 gttgtaaaac atttttccgt cgaacaatta tcacgggaag aaaattcaca tgtgctaaac 240
 aaaagaaatg tatggatgga acagaaccag ttggtaaagt tcttttaatt ttaagggtaa 300
 tgtcttattt tttcagatat gtcgaagcgt cttgcccag cttgtcgctt tgcaaaatgt 360
 gtggaagtgt gaatgaatcc aatggcaatt caagcagaag ttaaaactga tgaaggcaaa 420
 gttttaagaa gtgaagtatt gaatcaaaga gaatcgctag gggttgtttc atctgtaagt 480
 tgtttttttt ttaaaatata tctattaaat ttcagttgat agtaaccgag gaagatttgt 540
 taagcagaat gattgaaaaa ttaacatttg tggaaagtaa ggttgaacca ttacatcggt 600
 ctggaatgcc tcctggttac agagtaagtt aatgatTTTT ttttaatttg caagaacttt 660

aattgatttc	aggacataag	aaaattggaa	gaaattcttg	attcaaaacc	agtacttggt	720
gttacggata	ttcccaattt	ga'aattctgc	ccttcaccat	gtgcgaatga	aaaaagggtga	780
gttttttatt	aaaatgagat	tgaaaatttt	cccacttctt	tagaccaaga	cgacatttca	840
ccaactacgt	tcacacaagt	tacqtggctt	ctcttgaatc	ttcaaaaatg	tttgaatttt	900
cttcacaact	tgattttggaa	tccagaattc	ttttaatgaa	acacgcgcaca	ctaatttggt	960
cgaatatgat	gaatgcattc	ttttcgatta	acgaaatgaa	atcagatatt	cttaggcac	1020
cggatgggag	catgtcagga	catatgagaa	aattcgatcg	agagaatgat	gtcatgacag	1080
atcatgtcaa	attgatacag	aagacattga	tcactttcct	aaatcataaa	gtcgacaaaa	1140
ttgaatatct	tttattcaaa	gctattatgc	tttgcaatcc	aggtattttc	cttgccgtca	1200
tatatataat	ttaaacgatt	acttgaaatt	tcagctgtgc	ccggattaga	ttcctggaa	1260
caagaaatca	tcgaaaagga	acgaaatcaa	tacgttaaag	ctcttctaaa	ttattgtctc	1320
cttcaacatg	gaaaacttca	aggcccaaca	agattcgcaa	ctattctcgc	aatggctcca	1380
attatagaga	atcaatcgaa	aaatcaaaaa	gatttccatg	tgtacattaa	agcaaagcat	1440
ttccagaagc	acaaaagat	gggaacaacg	ttcagaaaa	gtatcagttg	catgtttgat	1500
caaattatgg	aaacataaat	aatcaatgta	aataaatgac	aatgtttttg	tttaactggt	1560
aatataatct	gttcatcctc	tgtataatat	tctacctcaa	aatacaaaaa	taaacatctc	1620
ataatttatt	cccctaattc	aaactacaaa	gcatgatgga	aactctatac	attagtcggt	1680
tgtcgtttct	tgtggacgtt	tcatgtggtg	ctgaagttag	gaatttgggc	gatttaccta	1740
cttgatgtaa	cttctaagtt	taacctgaca	tttgcaattt	taagaatttt	ccaagcatta	1800
taaaaaaaat	agttcacact	tttcatgtgc	gagtagatga	taattatgca	agttcaatct	1860
tgacacaggt	taaacctact	agttcacttc	tgatttttat	atattaaact	atcaatgtct	1920
tatttttgaa	gatttttcgt	gtgacagagg	tttcaaaaaa	aagagatatt	caatcagaat	1980
tctaaaagac	aggttgtgaa	acttcaatat	tttcgaatat	accgcacttg	tagtgaattc	2040
agaatataat	tttcaattta	aaadaacaaa	tttttgaagg	tacgcacqaa	tcacgcactc	2100
gaaaatttaa	gaactacac	atatatcagc	aattgatctt	caaaattata	acgacaaaac	2160
tgaaacttcaa	actttcaaaa	attccagaaa	ttaaaacgac	tgaagtgaft	cttgacagagt	2220
gttgactcaa	agtgaccaat	tggagaccat	acttgттаат	tctttcaaaa	acattttgat	2280
cgtcgcctgt	tacgtataca	aaagaagagt	actgattcat	caaatttggg	atataatccc	2340
accacatgt	aagtgtcgat	agttgttgaa	tcggctgaaa	actttttccg	aaacagaaag	2400
gagaaagaac	aaacatttga	tagaggcgat	agtcacatca	ttttccactc	tttttctcta	2460
ccatcctatc	tctaactcgt	ctagtcttca	ctaaatcaac	cgtttttata	ttccataaat	2520
tgatccatta	atgtcttcga	aggataggta	gataatattt	tcttgcgttt	attttatttt	2580
tccataatat	tccagtgatt	ctccaccaga	aacttttgaa	gtgtgcatct	ctccaccaga	2640
aatacacaaa	aaagaaataa	tgactccacc	acagaaatca	aaaccaaaaa	gaaactcgacg	2700
gaaaatttac	aaaattcttc	cacaaaatca	atgtccatcg	gtgtgtcaaa	tttgtagaaa	2760
tccagcaatt	ggatatcatt	atgaggtaaa	ctacagtata	gtttgtttta	ttgttagaag	2820
tagtgtaaag	gttacttaga	gctactttta	agattacgat	tcgatgagtt	attcgattta	2880
ctacatgaac	gctaaactga	gaatatgact	ttaaaaaatg	ttcaactttt	cggttttgag	2940
aaaaatcaaa	aggcggacac	gaaacacgat	tccacaattt	aagttttcaa	tactattctg	3000
ccttgaaaaa	tttatttggt	agaaaaagct	acatacaaac	ccaaaaatca	gtataaaaaat	3060
taagcactct	aaaagagtct	cacgttttag	gtaccatctt	gcaatggatg	caaaaacattc	3120
ttccgtcgta	caattatcac	tggccgtaaa	tttaaatggt	tcaaaagttag	taattgtctt	3180
gatggtaatg	acgttaatcg	tgagttcaat	aagagtctta	atgaaatttt	tttttgatat	3240
ttgcgttttt	gattttctgcc	ttttgatcgg	aaacacattt	ccattcaaat	tttacacaca	3300
ttttccagat	acatcgaagc	gtgtttgtcg	agcatgtcga	tttgaaaaat	gtgtacaagc	3360
gggaatgaat	ccaatggcaa	ttcaagcaga	agcaaagact	gatgaaggag	agaacttaa	3420
gaaattgatc	gctaaaaagt	ttgaaaacgg	agaaaaatta	aatgatggaa	ccgtattttt	3480
taatgtacat	gacagattga	atcaaatact	tggaaagtgt	ataaaaaattg	aaacaaaatt	3540
ggaaaaagtt	catgataatg	gaatgccaat	gggatttttg	gatcaaagag	atttaagcac	3600
tgcgctctct	tcaaaagtta	tctataacaa	tatggaaata	ccgtcaatga	gctatactcc	3660
tgttaaaaat	tcaaaaaaca	ccggatttgt	agtttttttc	aaaagaaaaat	tccaaaaatt	3720
tcaaaaaaatt	ctgaagtaat	gtttaaatcg	ttttatgagc	attgaagctt	gcctgattgg	3780
ccaaaaataa	acaaaaagct	ccaggcaact	gtaaaaataa	ataaaataaag	attataattg	3840
acaaacagta	tttttacaaa	ttgttgaaaa	ataagttctg	tgcgttttca	ggccaaaacg	3900
aaggagtcca	aactttgttc	attcaagttg	tcttgcatct	atcgaatatt	ccaaaacatt	3960
tgacttttca	agtgcaattg	acatttcaag	ttaaagtacgt	ttttggaaca	gattttatct	4020
gaacaatgac	cttattcgat	tatcttttga	gattatagac	tagcaccatt	tttgcaagta	4080
ttcaaaaatta	atatttagact	tatttcagat	aattctcctt	aaaaatactg	cactatcttg	4140
tgcaaacttg	acaaatgctt	atacaacatt	cagaaagctg	aaatcagata	cactttttata	4200
cccagatggg	agtattttatg	ggccaccacg	acggaaaaat	ggtcctttga	ttgaaaaaca	4260
acgatcattt	ttacaaaata	cgttgatatc	atttatgact	aacaatgttg	ataaaaactga	4320

atatatttta	ctaaaagcaa	tagtactatg	taatccaggt	atgtcataag	taccaaataa	4380
ttaattttca	aataaatatt	acagcaataa	tcgatctacc	ttatgccgat	tcaaaacata	4440
ttcaaagaga	gcgagaggtt	tacgctcagt	gtttgttccg	atattgtcta	ctacaacatg	4500
gtacattaca	cgggectgca	aggttttctg	ctttattatc	aattttcaat	gtgttgga	4560
atcaacagaa	ggaacaaaag	gattactatt	tatacatcaa	acttatacat	agtcaaaaaa	4620
acaaggatcc	tgagggtttta	aagaagaaat	gtattagtgt	aatttatgat	cagattatgg	4680
attagaaaaa	gaaatctaata	atgatagtga	ataaatagat	tgatgaataa	aaataatttt	4740
gaaatgtaca	tgtttacatt	tttatttgag	tacaattttgt	ttcttgaaac	catattaacc	4800
atgaagtcca	tcaattttat	cttacaataa	aaaccactcc	gacttgaaaa	ttaagtgtgg	4860
aatttgccag	tacttcagct	tcctttaaaa	aaaaattcaa	aatgtttaat	tcttttccca	4920
attttctatg	cggattttca	gaacttcaaa	ctattgggaa	cattttgaga	caacccccac	4980
agcttaaaaag	ttttcaaaat	tgtttctccg	ataaaatgca	catataccta	acttttgaat	5040
ataaccatt	ttctctctga	gttttatcgt	gatttctgtg	acgaaacact	gaatgcgtac	5100
tgcttttatc	ccgatttttc	attgatctcc	attaaaaaca	tcacagtttt	gttggtatta	5160
tctggtttac	aaagagtcac	attcagctgt	acatacaagg	gatctaatac	caacacatta	5220
tgtgcacacg	agctttgtat	cgcaacaat	acatcgcggc	taaaaatttt	cgttttctct	5280
actttctcat	tcattttatac	ttcatgcaac	gtgtgtagat	gagggaagtc	agacgcagac	5340
ggtacgtaga	ctccaaagaa	acaaagatgc	gttgatcggg	gcagacgcca	gtcctctctc	5400
taccgttcta	tggtttctgc	gattttatta	ttccaccttt	tgtgaagttc	gatcacttta	5460
ctttcttg	ttcaatatca	gaaagatgct	tcctgacttt	attacagtac	catcaacaag	5520
tgaaaagtat	gtttaatctc	cagaattggt	ttgtttgcat	agatgatggg	aaatcaataa	5580
cttctgttc	atgtgacg	cggtatcata	tttctcttct	tttttcagat	gtctttctcc	5640
ggaattgttg	ctaccaataa	tggtatgagaa	cttgggaagat	tcaaaaccaa	gcgtgagttt	5700
ttctattctg	gaaaaatata	ctttctaaaa	aaatatttaa	aaacaatcaa	aaagttaaag	5760
tattaattat	atatctcaaa	acgatataat	tcaaaaacgt	tatattcgaa	tttcagctct	5820
taaaatatca	aaaggcgatg	atgagccgag	gaacatgtcc	atcaaattgt	aaagtctgta	5880
ggcactcagc	tactggatat	cactatgatg	ttccatcttg	taatggatgc	aaaacatttt	5940
ttcgacgatc	aattctggat	ggaagaaaat	atacatgcct	gaagatgaga	aatgttttga	6000
gtggaactga	acctgttggg	atgctttaat	tttctcacta	attcaagatt	tcactatccc	6060
tatcatatac	ttcaaaccat	ttcagacttg	tctaggcgca	tgtgtcgagc	atgccgattt	6120
gagaaatgtg	tagaagctgg	aatgaatcca	tctgcaatac	aagccgacat	gaaaaccact	6180
gacggtgagc	tggttgagaa	agaaattatg	ataaagcaga	agacagccgt	tgattttttg	6240
aatactccac	aggtatatta	caaaaataat	tgcaacaatg	ataatcactg	attatttcag	6300
gttattatga	gttttgaa	taaagtgcaa	ggaataattg	gcaagctgac	agtaatggaa	6360
ttgaagatag	aaccattgta	tacaggagga	ttgcctccag	gcaatagagt	atgatgcttt	6420
tcctgtaaac	tggttaacttt	ataattcagg	atatcaggaa	attggacgag	ttgattgatg	6480
ctccactgat	tctcagctat	gacgaaattc	caaattttaa	gtattgtccg	agtgtagacg	6540
aattgaccgg	agagtggatt	tttattcaga	aaaatcaatg	attccataaa	attatttttc	6600
agaataaagc	ccagtggagc	ttgctacata	cattgcccgt	acctagctag	tatagaatat	6660
tcaaaaatgt	ttgactttgc	tcataaaata	gatgttgcta	gcaaagcaac	tctgataaaa	6720
catgcaacaa	tatatgtcgc	agataatcat	acagctttct	tctcttacta	tcaacgtaaa	6780
tccgatcgac	tgatttcatcc	aaatggaatg	tttgaggagc	caccaaataa	cagggtgaaac	6840
acttcagtaa	atatgttttt	gaaaaatgat	tttgcttgag	atatggtgaa	gccggtacaa	6900
aatatcaagc	ttcaatgcaa	agaactctgg	caacagtgtc	ccgacatgaa	cttaatcgaa	6960
ttgaatatat	gctactcaaa	gctattgtgc	tttgcaatcc	aggatgacc	aaaaattttg	7020
aatcaaaaac	acaaaaacat	caatttcaat	acaggagtcc	atgaaaataa	tatcttttaa	7080
taacacgatt	tgattttccag	cagtctctag	tctttctatt	tcggtacaac	aaataatttg	7140
aaaagaacgt	gaagaatatg	ttcgaactct	acttacatac	tgtcttctca	attatggttc	7200
tgttcatgga	ccatcaagat	tttctgccct	tcttgctatt	atgtcagttc	ttgaaagtca	7260
acagaagaat	gcaaaggact	ttcattttat	ggctaaagcg	acaatactga	aagatgcagt	7320
tagatacaca	cgaattagca	atctttacga	acaaatcatg	gagtcgtaaa	gattagaaat	7380
tgttgaaatt	aggatatttg	ggtactcgaa	gtctttattt	ttattacttt	caatttatgt	7440
tatatattcta	tttatatata	catttgagaa	atacaaagaa	atcagagaaa	atcatcagaa	7500
ttgttgcat	ggttctagcg	ggtattgtaa	ataaaatgtt	atttttccgt	ttcgaatctt	7560
ctctacataa	agcttttcga	tcagcacata	aagtgagaat	agaagtgttg	attaaattac	7620
aacctgaatc	cgttatccgc	aatcagttta	accgcacgt	aaagacaatc	agcagtgaga	7680
tcaggagtca	cgggcatctt	ttcgaaatct	ggtctccatt	taccagttct	cacttgaaca	7740
cctcgcatte	cacaagcttg	agcaccgcct	acgtccgaca	tcaaatcgct	gcccaaccatt	7800
acaatctgga	aattagaaaa	tgaagtcaca	aagattttgt	tccaacataa	gaatggaaat	7860
tgaaattccg	atgtataatt	gttttactta	cttcttctgg	tttcataccc	aatgcattca	7920
ttccttgctc	gaaatagaac	cggcttggtt	ttccaatgtt	caaaacctca	caattagttg	7980

agaacttgag	agcagcagca	aaagcgccga	catcaatgca	cggaccatca	actcgttgga	8040
agaactttct	aaaatagapa	agggaatgt	gttagagaga	acaggaaaca	attattaccc	8100
attgcccatt	gttatgagaa	gcggtttagg	catgtcgatg	agaattcgga	aagctctgtt	8160
gattcgatcg	aatgagaaac	cttqttcaac	ttcaccatc	acaacacaat	tcggagatga	8220
ggtgtcgatt	ccatcgaagt	attcaagcac	atcttgagag	gaaaaagaac	tgattgttgt	8280
cagagaatta	aagaatcaaa	taccgtctct	gacaaacaga	tgtgggcgaa	gcttgttctc	8340
tcgacaatat	tgtgcaacca	ctggagctgg	agttatgacg	tcttcctctc	gaacactgaa	8400
aataaaaaact	attataattc	caataagtag	tgggtgattt	gaaagctgcc	tatttgacat	8460
tattaatfff	tacattatfc	ggtaacactt	tttttctaaa	tttgttctat	cagacttcag	8520
taggacattt	tgggtggttt	tgaattgatg	actgatcaac	agaaatattt	tctgaatctc	8580
aaaaaatcaa	tggctgggag	agtgaatttt	tttttctgat	ttcttccatc	ggaaaacatg	8640
caaaaattgg	gtaaaatttg	atttctacaa	daaccgtgaa	ataactctca	aacttacttg	8700
attccgagtc	tttgaagtct	ccgagcaaca	tttctgtttg	aatttctctt	tgcatgtgta	8760
agaaatttta	ccttegagtg	ttgatacaaa	ctgtaaatat	tttcaatatt	atgtatctat	8820
ttttgcattt	tacaatatft	ttcttatcta	atgcaccgtt	gttctaggtg	cacatgcacg	8880
caaatgttta	ccttgagggc	atfttttata	aacagttttc	caaaaatttg	tggaattttt	8940
ttcggaattga	gaacaabata	caggacatgt	aaaacgattt	gacataaaaag	tgataaactt	9000
gaaaaatgag	gatacatgat	ttgtatctac	cattttttatt	aacaaaaaag	agttgaagtt	9060
tttattagct	agaaaaatca	ttaaaaaact	tacaaattaa	ccgcttcagc	tgatttttga	9120
acagcaactc	catcggattt	ataaatacta	ttatacaaaa	ctcccgttat	atcgagcagg	9180
aaccatttaa	ccgctcttcc	atttgacatt	cttcgaccct	aaagtttcaa	tatgaaattt	9240
cataatcctg	aaaccaagac	caaagttcag	gaaaccagca	aaagctataa	catgaaacaa	9300
aaacagaaaa	gatagaatga	aaatcttgag	atagctatgg	aaaaggaaga	gaggaaaggg	9360
tctttctctc	tttctttcca	tgggggggca	atatgggcag	aaaggggtcg	tttagatgtt	9420
cttctctctg	ctcactgcat	gtgcaccag	tacgtcttct	tcacttctct	cacctctctc	9480
actcctttct	tgttccacta	tctcttcaat	acacatgtcc	agctgcctat	tggaacttct	9540
gttttcaagg	tgttgtcaca	acttttgagc	tatgaaacta	ctttgatgct	aattagatat	9600
gatttagtaa	tatgtatfff	gtttcatagc	tcacacagaa	aatgtatatc	atgagaaaca	9660
tcagagagat	ttagcaatgg	atccgtaatg	tatctttaat	gaactcaaga	tattcatatt	9720
ttaaaacgtg	tttacctca	aaataaatga	atattgtaag	aaaacactga	atcgcatacc	9780
caaaagcctc	caaaacttca	aaagatgtag	agtaaccata	caaaacaata	ggaagagttc	9840
gaaatgtgca	gtcacgcgcg	catttcttgt	gtctctaagg	tgtgaattac	gcaaagacag	9900
agacggcaga	ttgcggtatg	tcgggtcgaga	atgatgaatt	acgatcaaga	caagaaaaga	9960
agaagaagaa	ggaaagacga	accagagatt	ttcgactgca	gtagctgagg	aacaattgta	10020
gaagaatact	cggcgtcact	ttataaataa	atgaaacaaa	atcttgatta	aaagagaagc	10080
tagaagggtca	cgtaataact	atgaggttat	gcacttgaaa	ttatgcaata	tttcaaattg	10140
cacagaaaaa	accaacaggt	aataaaatca	atcaagtctg	aaaatgttgc	tctagaaatg	10200
tgtatcgtac	aatgaagcgt	gtatctctga	atcaaaaacg	tttagactgt	gacaattctt	10260
tccggaatcc	aataagaaca	aagtgtggaa	aagggatgca	ccgagctgaa	aatttattga	10320
actgtagttt	gaggaactca	atcaagacga	acaaaacatc	attcatgcga	agaaacatca	10380
tttaatttta	gaaactataa	aaaaaagggt	aaaacattta	actggcattc	agagacaatt	10440
cttaaaactt	tgatggcgca	aaacaacaag	acgtcatcca	cagaagaaac	tgtgccctac	10500
atcttcaaat	agatcggaca	tcacttatct	acaagtattc	tttctttttc	attctcttgt	10560
ccctctccat	cagccctcct	tgatagtctt	ttttcacatt	ctcaccttct	tttttgctgt	10620
ctatttactt	ttatgattgc	cgcaccaata	aatatctcta	ctcactttca	aactccttat	10680
cgatttaatc	catatgtact	atacttttaa	agacagatta	tgcgtgaggc	gaatgtgaca	10740
tagtctagaa	ttccggaaat	tgtgcgcaaa	attcttagaa	acatatgtag	ttgagtcgca	10800
tttttcgaat	ccagttttta	tttcagcggc	tgttttattt	cagcggctgt	tttatgactg	10860
catttttaca	agctgcta	tgcgaatttc	cacgataacg	gtttggttgg	cagtggttcc	10920
ccaataaatt	tgtaagcatt	ttacataatg	caccaatttt	tccctccaca	catcatttcc	10980
gaatttgatt	tttttttctt	aaattacttg	aacatattat	cgaaaaaaat	tgaaactgtt	11040
cccatactca	caaaacatca	ttcaatagta	ttagattcag	aaaactgcct	acaaattgat	11100
tttttaacca	ttgggttttca	tgaactatct	ttcaacaatt	cagatgaact	ttttttaaaa	11160
ttcaaagtgc	tattatgtta	ttacaattag	aagaaattaa	aaattcaaat	tatacaattt	11220
aagtttggtc	ctatcacggg	ttttaaatga	atttcttgaa	aacgcgcggc	aaaagcttag	11280
caaactgtct	attgaactta	cccttactat	taatcattga	aattttcaaa	tgcgatatatt	11340
taaaaaaaaca	gaacggtag	gatttctcat	ataagacagt	tccttaatat	ctcgtagcaa	11400
aaaaatcgac	ccatctcttt	gccccttgct	tcatttctgc	cgctttcttc	gtgccaggaa	11460
ggtaggcatt	ctcgtgaacc	gacggtctga	gttttagtgt	tttctcttta	aaatgttcga	11520
ttattatagg	tcgccagaca	taaatattat	ttgttaatat	ttcatggtgt	cgaaaccaca	11580
atcttttttc	tttctacagg	tggtttaata	ttcgtttctt	ttgtaaattt	tcccatcggc	11640

aggaattggc	ataacctgat	tttttacagt	agaaacacat	atctagaaga	ttaagatcctt	11700
caattgtcca	tctcgtttta	aactgtcatt	tctccggttc	tttttggttt	ttggtggcct	11760
tttaccgggt	ttgttacgtt	ttcccgaat	gatataaatt	tttgagcttg	tcacttaatt	11820
tatcaagcgg	cgaccaaaga	caaatacaca	ccgcagaaga	aaagtttgta	aagagaaatg	11880
gaagttagct	atatgtacat	atataaaatg	gaatgaatag	tcgcatcaca	ttgacgggtga	11940
aagtacatgc	ggtgatgcct	tcgaggacat	acggacacac	tttccttctc	caatttcac	12000
gattaatttg	tattgttata	ttctgaattc	caggttttca	agacgaacat	tgtgaatcat	12060
tttcatcaca	ttgttctggt	acaaataaaa	tgtatttcca	agtgttcccg	cacatttaca	12120
tttttgcttc	actacgtgtt	tagcacgtta	tgcgataaga	ggagaagtgg	ctgtaaaatg	12180
gtttcgact	cccttttttc	ttcccacaa	tcattgcacg	gtttctatga	ccttttcgaa	12240
ctacattcct	tcttttttca	tttgagggtt	tgatactgat	atagtttgta	gggtaccgc	12300
ttattgtaac	tetaactgtt	atatccattg	tcattgattc	atztatctgg	ggggtattcc	12360
cttctatttg	gcgaagtcta	catttctcgt	aaactgtaca	aaataacgtt	taaaagtctt	12420
ttgaatactg	aaaataataa	aacggtttga	acaatccaac	ttattttcca	attttcaaaa	12480
actagccgat	cattttttgtg	aaataaaact	attttctgct	tagtatcgaa	taaaacatct	12540
taacggtaaa	aatgagtttt	ttaaaacact	tcttttaaac	cttttttgga	tttatttgaa	12600
ttttggcttt	gaaatagtct	ttattccgga	aatttcatat	aatctcaagg	gcaaatacat	12660
tttgaaccgt	gttcaaagtt	atttaaaaat	gcatttcggt	tgaatcactg	agatttccaa	12720
aaattcaatg	agtcataata	catggagcaa	caaactata	agaatgtcca	cttgcttttg	12780
ttgttctttt	gttttcaaac	gttttgata	taggtctgta	ttttggccga	caacaacttc	12840
agaatgtatt	ttatgtttga	gttttaaat	tgttactgaa	aacaaataaa	atgacagatt	12900
aatatatctc	ttggtaaaaa	ttaatatatt	ttgttactac	tttctgttcc	tgaaaagcct	12960
gaaaaatcag	accatttttt	gcagtgtcct	tgtaaatgat	aaaaaccgac	atgcacactg	13020
accatcaaac	cgattattat	gaattaataa	tttaatccga	cagtttctctg	tttctgttca	13080
tttcatctca	cgaatgcttc	aattttcatt	tttttttcc	gtgtctctta	atacgtttaa	13140
tttattacga	gttcaagtaa	caacgtttca	atgaactctc	atthagtttg	aaaattaaaa	13200
tattttaagtt	ctaacatttt	gaagagtga	cgatggtgag	aaacaaagtg	gcaccagttg	13260
aagatggagc	caatattcaa	agaaacttcg	agcctccgcc	accatataca	actccaacgg	13320
attcacctga	agacaagatt	cgatcaaatt	caactgcaac	aaccgcatct	caaccagaat	13380
ttcaaggatg	ttggactatt	gttgtagttg	ccattttatt	catcatcaat	cttctcaatt	13440
atatggatcg	gtatacgatt	gcaggttagt	tgacagtaac	tatttttacct	ccttggtaca	13500
atatccaac	cacaattcat	gtaaccatgt	gacagttttc	cgtgcaataa	aatttagtag	13560
agcctccaaa	actataaaca	acaaaaggta	gaaagatgaa	aatggccgga	aaacgacatg	13620
agagagtatc	ttgtttgaac	ttgtcatcat	tagaagataa	gaattttacgt	aatgggtgag	13680
tacaataact	atgggacaaa	catcattttt	tgatctaaac	ggagcattaa	aaaatatgtg	13740
catgtcaaat	ttgaattagt	atgtgtggaa	cttttcctac	tgcaataata	taacgttttc	13800
aacaaaaaaa	acattaaaaa	cattattcgt	agaggaaatt	agcattttaa	aaccaataaa	13860
attttttagga	gtccttaatg	atgtacaaac	ctactacaat	attagtgatg	cttgggctgg	13920
actgattcaa	acaacgttca	tggtgttctt	cattattttc	tctccaattt	gcggtttcct	13980
aggagatcgt	tacaatcgaa	aatggatttt	tgttgttggg	attgctgatt	gggtgtctgc	14040
agtgtttgct	tccactttta	ttccatctaa	gttgagtttt	tgactctaatt	catatctattt	14100
cttcaatcaa	cttcattttc	agcaattctg	gttggttctt	ttattccgcg	gaattgttgg	14160
aatcgagaga	gcattcttatg	caattatctc	cccaactgtt	attgctgata	tgttcactgg	14220
agtgtctcgt	tccagaatgc	ttatggtctt	ctactttgcg	attccctttg	gatgtggtct	14280
gggattttgta	gttggtatctg	ctgtggctag	ttggacagga	cattggcaat	ggggagtacg	14340
agtgtactgga	gttctcggaa	tcgtttgtct	ccttgcttatt	atcgtgtttg	ttagagaacc	14400
agaacgtgga	aaagctgaac	gtgaaaaggg	agaaattgct	gcattctacag	aagcaacaag	14460
ttattttggat	gatatgaagg	atttgctttc	aaagttagtt	tctttcaatt	ttatttaaaa	14520
agcaaccttt	catattttcag	tgcaacatat	gtgacaagca	gtctcggtta	tactgcaact	14580
gtttttcatgg	tcggaactct	cgcttggtgg	gctccaatta	caattcaata	tgagatttct	14640
gctagaagga	acggaacaat	tacagaagat	cagaaagcaa	acatttaactt	ggtattcgggt	14700
gcactcactt	gtgttggtgg	cgttcttggt	gttgctattg	gtacattggt	ttcaaatgta	14760
aatattgtcg	ttgaagatta	gaaaaatata	catctttttt	ttagatgtgg	tctcgtggag	14820
ttggtccttt	taagcacatt	caaacagttc	gtgcagatgc	tcttgctctgt	gctatcggag	14880
cagcaatttg	cattccaaca	ctcattcttg	ctattcaaaa	cattgagagc	aacatgaatt	14940
ttgcttggtg	gtgtaaaatt	tggtactatga	aacatttcaa	attcgctttt	ttttagggaa	15000
tggtatttat	ttgcattgtc	gccagtagtt	ttactgggc	tacaaatgtc	gatctgttac	15060
tggttaagttt	cactcgatat	tgaacaaaat	tcccaataaa	atgttttttg	cagagtgtgg	15120
tagttcctca	gagaagaagt	tcagcctcat	catggcaaat	tttaatttct	catatgtttg	15180
gagatgcttc	tggtccttac	attctgggat	tgatatcaga	tgcaatcaga	ggcaatgaag	15240
atacggcaca	ggctcattat	aaatcattgg	tcacctcatt	ttggttgtgt	gttggaaactt	15300

tgggtgtgtc	agttatttcta	tttggaaatat	cagcaataaac	agttgtaaaa	gataaaagcca	15360
gattcaacga	aattatgtgt	aagtttcaga	ctttttatag	cttggcactg	aatcaaaaata	15420
ttcattttgca	gtggcacaag	ctaacaagga	caacacatca	agcgggacac	ttccaattga	15480
agacagaaac	acagaagacg	aaactggttc	cgaagttcaa	catatgtaat	tctaattttt	15540
tagtaaaagct	ggttgctcaa	ccctctctcc	cctacatatt	acattttttot	cacagtttat	15600
tgcgggtctg	ctcttttttt	tgtatttttg	tatccccat	cagtcaatca	ttcgcataaa	15660
ttttgcgttt	ttattgccaa	atcttttgta	catttcggaa	tattcgtgct	gtcagagaaa	15720
atcaatgata	gcagttaacg	tgattcttta	tgcttttcat	agtgtttttg	aaaaatatga	15780
aaacaagttt	ctacttgga	cttttttatg	gagcctcctt	gttgtgtggc	ggacaaatga	15840
caacagaagc	tgggtaaaaa	gtaagtttta	ttcaaaacat	actatacttc	gccttcccac	15900
ctctcgcccg	gaaatgtttt	cggaagagta	tagtaaaaggt	cagttgaagt	gttaacatgg	15960
taacaaaagt	ggcatacaaa	actaccggtc	aggcgcgagg	aggcggaata	agaggcgagc	16020
atgaatcacc	aattgtggaac	gccgttattt	tggtttattg	tagaaatatt	actttcgaa	16080
aaaatcgtga	cgccaaacaa	aaataacttt	tgcgttagct	tgcagtttca	atttttaaaa	16140
ttttccatgg	ttgtccaaat	ttctgcatag	caaagtactt	gtcacgtttt	tcaacataag	16200
aacagttttc	agaaaccaac	tttgattcat	ttttcattta	cataatttca	gatacaattg	16260
aaatctgcaa	tattgtaaga	ttcatatcag	ttccattatt	tcattcatct	ccagtttaatt	16320
ggctttattt	atttactaca	tcttcagcat	tttcttttat	tgctcaaagt	tggtttctaa	16380
aaaatcaaaa	gtttttaaac	tgaccaattc	agaaaaagaa	tttactggat	tcacattaat	16440
gatcgctctt	catttgattc	aatattttct	catcacagga	ttaagtgcga	tgacataattg	16500
gaatttagga	aaacaattat	ttaattttatt	tgtttggcca	gatgattctc	gtcaactaat	16560
cattcagttg	gcaatgattt	tgtctactgc	aactatgttc	tgtcacgtat	ttatcagtat	16620
ttggtcacaag	ctttcagaga	ttatcaatat	aagtattgtt	tttttacatt	gtttctctga	16680
aacagtatta	tatatatttg	cttcagattg	gtttctagtt	tctaccatcg	ttcatcctgc	16740
tattagtttc	tacatctttc	cgagatttca	acatgttcat	atgttgactg	tttggttttc	16800
ttttcttatg	atttcagtta	aatttaataca	gaagatgttc	attaaagttc	tcggagtctt	16860
tttcattggt	ctatactcca	gccttttgga	atcctgggaa	ttatattttt	tggttttgaa	16920
gtcacttttc	aaaataaaa	atttattcta	acaagtgttt	acgtagtaag	cattaactta	16980
aatatggtga	aatggaaaa	atggtgctta	aaaaacatca	aatgaacac	agactaccga	17040
atltgataat	gacacttcgg	taatccttct	caaaaatttt	tattactgtt	ttaaatgtta	17100
gaataaaaag	ctcaaaaaca	aacctccaac	tattagcagc	cggaagaagt	aattaaattg	17160
aaacaaattt	aggcccgctg	caagcctttt	ctttataaat	tttgaacaat	tataaattca	17220
gtctttgaaa	gccaatttga	gtttttattaa	gcagtatata	tttttagata	gtcggcattt	17280
gaatctatca	acaattttta	tcatcatcat	catgtttatc	atcaacagta	cctccactgt	17340
ttcgtacaat	atcatcatat	cgtaaataag	ttcttctgtc	agcatttgat	ggacctggac	17400
gaacacaatc	cgcatatcga	agacgatgaa	atcttgtagc	atagggttga	ggatctctga	17460
aattgttcca	attttcttcc	agtgttattc	cgaaagaatc	gtaccggtaa	tgtaaaatcc	17520
atltctcaaa	aattccttca	aaacaagctt	cagtgtcaga	atltagaaga	agtatctgaa	17580
tcaaacaaac	ctaatacagat	gaaactttta	acttcatacc	caattaacac	accaagattg	17640
gcttaaaactg	aatccaggcc	accagaatag	tcccatattg	acaatatcgg	atgttttaaa	17700
gaaaatcttg	aatgatcac	gagatccttg	ttcaaaatcg	tttcttctcg	ggttatctgag	17760
atgaaaagtt	tcagaatcat	gtccagattc	atagaatact	ttcaaatatg	atgcggcgctc	17820
tgttccagca	ccatctttgt	cacaagtctc	tatgactaca	tcccagagac	gacctgcac	17880
tacaaaactg	aataattgcc	acttatttga	atttgaatac	aacaattttc	agaaaagttt	17940
tattatgatt	gttggttcatt	tcagcattgt	atgcgtgatt	ttaaacgatt	ttcttgattg	18000
acacattagt	tgaactcaaa	gttaagatta	aaaataatat	taaactacat	cattgttcac	18060
aacttcctgg	aaagttttag	tgaggttcat	attcgaatat	tctcttcaac	cccattttta	18120
aaagttgact	tgtgttttaa	ttgtttaccc	tttctgatca	tttctatcat	tgaaattgga	18180
agtcatttca	cttttcttcc	cacagctcat	agttacaaaa	acaagagtga	ctgaaatttc	18240
aaaaactact	tgtggtttca	gtttttattt	tattttatga	atccagttag	tcgaaaaaaa	18300
aaacaatcaaa	tttaattccat	gcaacccaaa	atatttttaa	gttcttggtta	tttatgcagt	18360
ttttaaaact	gagaatcaaa	agttaaaata	attgttttaa	acttacgatg	aacaacagtt	18420
tcattagttg	gagaaaattg	tgcaaaaacg	atagtaacta	gtaagcagta	tattaaaaat	18480
gagtggaaaca	ttttggggag	ttggtgaaaa	aattatagga	agagatgaaa	gacactttgt	18540
gttctggaat	cttctcctga	cgacaactaa	aaatccatca	gcaattgaaa	tgatcgattg	18600
atgattgatg	gatgagtgtg	atcatatgct	tgctcaaaat	actttcaaa	cgaaaaaaa	18660
acacaaggga	aacatgagtt	ctaaattttt	gatcggaaca	gatggtagaa	attgaaatgt	18720
ttggattcgt	catcggtaaa	atgaatatga	acattttatta	gtgcaataaa	acaatcaaaa	18780
tttaatatga	aacaagatga	aaaatatgac	ctgatttttt	gaaattattc	tggttctgaa	18840
aaaatatattg	tatgattata	caaatcgaaa	gatttcaaac	tctcaaatg	atgtctctcc	18900
ggaatctcgt	gtattttcac	gagtttcagt	tctttcaaaa	aacgtgccaa	atattttctt	18960

ataaaacaaa	tagtctttga	ctaaatcaaa	gattaccgtg	atatgagatc	ttaaaagtag	19020
cattggttat	aagctattga	tatgatataat	acgagatact	gtaattgtag	atthagttta	19080
gcttcatata	tctctaattt	aactttttta	gcctcgagtt	aagtatttta	catttgaaca	19140
attgtgtaat	caaaacgcaa	tcctgtacat	aattatittc	agcgctgtat	tttctctcat	19200
caatcaaaaa	atttaatttt	taaaaataat	ttgattttacc	gagcatctga	acagaagcaa	19260
ggacacgtga	gtttgctgat	tttgctgata	agaaatcaat	tatttaggtg	cattgggtgat	19320
cttcattctt	tttctaaact	tatcagaaaa	aatataattct	ctatttttaa	catttattgt	19380
gttctacatg	ttctacatgg	ttttgtatat	tcgtaacaat	tttttaagta	gaaaaaata	19440
tatactgata	atcagcaaaa	actagttttt	gttcacattt	tgatctcatc	acttttttat	19500
tatattcaac	tagagtttaa	aatattaatt	ttattgattc	ttatcgatat	ttttgactat	19560
tgattcccg	gttgatcgct	ttgtgataat	aaaatttttg	gaacaaatta	gccaatcggt	19620
cattttttgt	catcaattca	atatpatctg	acgattccctg	ttcctgttag	ttgcaaaatc	19680
atgaacaagt	atcacttttc	ctatatttta	tttcatattt	ttcccgtcat	gaaacgttat	19740
gttttgctc	atataaatac	aaaaatgaac	aaaaaagtgg	tgttcgactt	gaaaacaaga	19800
tttttaggac	agaaggtcgg	atgggaaaag	gttttcgttc	gacggacatc	acacaaaaaa	19860
gtatcacacc	gcacctcaft	ctattttactt	tgttgataga	atggaaaaac	attcgactga	19920
aatgaatac	agagtaagtt	tatttttcca	aaaatgtttg	aaataatgat	tcaattttta	19980
ggtgtataag	gaacgatggt	ttattctttt	ggcgactact	gcattgattt	gctcaaatat	20040
tatggtgagt	aatggaatat	tggtattttc	atatttagat	tggcgtgaat	tttgctaaga	20100
agtcgatgt	ttttccaaaa	ttcaaataca	ttacaaagta	gtgccaaaat	taaacatttt	20160
ttagattttt	tgtttttagct	gaaaagtttg	acgaggtgtt	cataaattta	aagagttttt	20220
gaaaacaaac	ctttaattat	gatgttcctg	catittcttac	aatttttgag	gtaatttaat	20280
agaagacgag	cctgccacaa	aacggtttta	ccagattttt	caaactttta	caattattta	20340
ctggatttaa	aataaaacat	attaatgttt	tttctgcagc	aatgggtgag	tttctcggct	20400
caaatcgacc	aaacaaatat	gtttttctgt	ggcccaaatg	aatataagaa	ttgttcagca	20460
gccttcctat	ctaatacaat	ctatcaggtt	agatttttca	attttgatcg	gtataattta	20520
ttatattatt	tttagatagt	tgcatgaatt	gtgagcatta	taggaatgta	ttttgcaaca	20580
gtttttggaa	cacttccaac	tgtatgtttc	ttcttcta	ttacagtttt	ctaatttaca	20640
cgtcttctag	ctccgtctct	cagcagtttt	caatattgtt	ggtgcagcaa	tccgactcat	20700
tgcgtcactt	ccaagccttc	aaaacttttt	ctggagacag	tttttaatga	ttattggaac	20760
tagcattgcc	gcggctgctc	aaatgtattt	cgctcgtctc	tcgaaaattg	cggaaagttg	20820
gtttatcca	agacatcgag	cttctgcaaa	tgttgcattg	tcaaatagtc	ttgagttagg	20880
agttgtactt	ggaactgttt	taccatcaat	tattgtacca	gcttctttta	caaaagacat	20940
tgtcagttcc	tggaactttt	ttcttttggt	tagttatgtc	tctgtcatga	ttaaacaact	21000
atattacaat	ttcagaattc	aataatagca	gttgatgta	taattccatt	atttttatta	21060
tttgttctat	gccgtcgatc	ggttccaaaa	acgccaccat	cagcctcatc	gcaacatgag	21120
tcaagtggtc	ctgtgacatc	tggtattttt	aaatgtttta	agtaagcccg	tacttagctg	21180
tttaatttta	attcttctga	tagttctcag	ggatcgtcag	tttctcattc	aaatatcgt	21240
ttattcaatt	aattttgcaa	ttgcaaatgg	cctcatttat	acttctaattg	ccatcaacta	21300
tagaggttat	aatttgaaaag	ggtacgtcta	gtaaggtttt	ctgctaggta	caagaaatcc	21360
tactctgtca	agtaggcagg	catttttaga	cctgcctagg	aggtttaaac	tcaagacagc	21420
cccttgcgta	taattatttt	gttcagggtat	ccaatagcaa	tgcgccagat	ggtctgcattg	21480
ttttcgcctt	atttcgtagg	agtaattgca	gatagaacaa	gaaagttcaa	agttagtttc	21540
aaaaaatcta	atcttatctg	atatccattt	ttctagttga	tcgcaataat	aaacgcactt	21600
gttgtagccg	tttggtgact	ggtacttcgt	ctggtaaggc	tttttcaaaa	taacattaaa	21660
tttattcatt	aatgctttta	gtatttaata	aaaacttata	cgggatggta	tgattcagta	21720
attgtttgta	cacttttaag	cattatcagt	aagttatttt	aattttgctt	atttagaata	21780
tattttcaaa	aattcagttg	cctgttgtgc	tatacatata	ccgattggaa	atgagatggg	21840
agttgagaca	acttaccag	ttcaagaatc	aatttctaca	ggagttttga	atactttcgg	21900
gtgagccttt	ggttcttttt	acggcgaaaa	aaagtgtgat	atacccgctt	ttttgggtct	21960
cgcagcgaat	ttgtatttcg	gtttgtgctt	ttctgatatt	aattttatac	gagttttatt	22020
caaaacgaaa	cgttaaaaaa	tttttcagtc	aagcttggtt	attctgctta	tatttttata	22080
tgatttcttt	acaagaatca	aattgggttt	acaagaacaa	ccgtaaagga	ggaagctggg	22140
aatgtaagtt	tcttttttaa	ttttaaaaca	atgtttcaac	taacgcggtc	acttacaagc	22200
ctgtgtagct	ttagttttct	actaattttc	aaataattaa	aagttttttt	tctaatagta	22260
aataaaaaata	taaataattac	gggaacacta	aatactgaaa	gtgcgtattg	cacaacatat	22320
ttgacgcgaa	aatgatattg	tagcgtagcg	aaaactacag	caacttttta	gatgactacc	22380
gtagaactgg	tttacggaac	tcgtgtttca	atcgataaaa	tattaaaaga	aaacacaaaa	22440
atgacaaaac	aataaaaaaa	aaacttttaa	aaaaattgaa	tattgttgct	acaattcgca	22500
caaatacttg	tttcgtaact	agagccccac	ggcccgagaa	gtggtacctg	tacgcaattt	22560
gtctaccgta	cacctggacg	tttgggcgcg	tttttctcaa	aaacggctgg	tccagttttt	22620

```

ttgtgatgca tataaaaaat gttcgaat aaattcaaaa ttttttggac cdaagctttt 22680
tccttaaaac gagcccaaac ctggctaaac tgcaattatc aatagagcgc gtttacactg 22740
atgtaccctt tgccgggctg tgagcccgta aaccgacacc agcactaagt acagaagtca 22800
ttaaataaat tgtttttaatg ttcgctacga tatatttccg ggtcaaatat gttgcgcaat 22860
acgcattttc aaaacgtagt gttcctgtta attatgtttt catttctgct tcaaaaaacta 22920
aatattttat attacagtag ccttgactt ctgggcagga atgtctattc taaactttat 22980
tgtggctgta atattcttga gaccccgata taaccgttta caadtggaag aagaagctca 23040
aaatacacgg gaagcaattc gagattcagt gtatacaatt tgtaagtaat gaactgcttt 23100
tatttatcaa tgatgtaaata aataaaaaaa tgaatttgct gaaaaaaata aagtcatgag 23160
aaacaggaaa tgcgatatga gacaatggtg aaaaaagatt gggactagtt gaaactgggt 23220
gcaaataaac tggatgaagtc tagaaacata atataaacat gtttggtggg caaaataatt 23280
tcaaattgag actatggtac atatgatttt ttaagtgaga aaatacaatt cctctaactt 23340
cttaagtaca aaatgctata aaatgcaaa gtaaatgcttg ctactatcgc atttcgatat 23400
ccgacgcagc cgttogatgt gtaacaggtt cctcattttc taatgccaat ttggcatttt 23460
ccagttctag acgacgatat ctgtaattga taaaattatt tattgagttt tgataaacgt 23520
ttataccgtg gtctcaaaap gaacagtgat aagatgacag atacaattga gagccacat 23580
gttacatcga gagcaactga aaaatattca tttacttcta ttttgaagat gataaaactt 23640
acatttccaa gattgtgcat agaacaagt gcaacttttct gtgtaatttt gtgcaaatgg 23700
aataatgaac atgaaaagcg atccgaatat aacaagaaca cctgatgacg ttgcttccat 23760
tactgggaat gtcgtctcaa caccaagttc cactccaate ggggaattgag gtattgaaaa 23820
ggcacctgaa gaaatatagg aattttgtta tacaattaaa tttaaaaaab ccttaccgag 23880
acatccacaa agagtgtaaa cgattatgct atcaaaagag ccagtcctgt gttgatttag 23940
gaactgcaag tgtagccttt tttagattaaa aaaatccata gaggaatttt agacatatgt 24000
acacaatttt tttcgaagac cctggtagaa gaattttaac tcaccattct caatgtaatg 24060
actgaacatg aaaatccaac agtacaaact ctgattattt cctaaatttt aaaattatat 24120
tttctttttc tgattcatac cgaccactca ccttgaactt gcgtgttttg tcggcaatat 24180
gaccagcgag cagtgatgtt aacgtgccaa cgattgcaca cacggcagtt ggatatctga 24240
aaaataaatg aattactcat atgcttctag aactctgcaa agttgttacc cagccatctc 24300
ataaccttga tctttgagtg gaccatcgag gaaaatcatt agactccaca gaagcgaaaa 24360
cgcaaatgca aagagagtca tttggatgaa aaattgagca ttactgaaat taaagtattt 24420
aaagtttttg gcgtgatact aaaaacctac aaaatacatt gaagaataga tttaaagaaa 24480
ccaatgttgt tctgatgagc agcagaagaa gctgatggtg gagtgggtgg cagtttagtt 24540
cgaacgaata gtgctagaac aaatgggaac aaggctagac attccattcc aagagtctga 24600
aaatatgaat tgtcaatgat aattaaattg agtactaaca aatgtgaaga acatccaact 24660
attcgaatca atagtcttat tgtgaccaa aagaatggat ggcacaattg ttccgagggc 24720
aacaccggca ggatttgctg aaattgatag gaattactta cgaacattaa aattccacta 24780
accaacaaaa gaaagtacat ttgcaattgc tctctgatca ccagggaacc agcattcagc 24840
aatttttgat ggaagaacta aaaagaatgc ttgtgttgaa gcagcgataa acgatcctgc 24900
atgaagaaga cattcacgaa caaatgtga ttttatgaaa ggaatcgagg caatcatccg 24960
aatcgatgct ccaattacat ttaaagttgt accaagaaga cc 25002

```

<210> 32

<211> 1237

<212> PRT

<213> Rattus norvegicus

<400> 32

```

Met Ala Arg Ala Lys Leu Pro Arg Ser Pro Ser Glu Gly Lys Ala Gly
 1           5           10          15
Pro Gly Asp Thr Pro Ala Gly Ser Ala Ala Pro Glu Glu Pro His Gly
          20          25          30
Leu Ser Pro Leu Leu Pro Thr Arg Gly Gly Gly Ser Val Gly Ser Asp
          35          40          45
Val Gly Gln Arg Leu His Val Glu Asp Phe Ser Leu Asp Ser Ser Leu
          50          55          60
Ser Gln Val Gln Val Glu Phe Tyr Val Asn Glu Asn Thr Phe Lys Glu
          65          70          75          80

```

Arg	Leu	Lys	Leu	Phe	Phe	Ile	Lys	Asn	Gln	Arg	Ser	Ser	Leu	Arg	Ile	
				85					90					95		
Arg	Leu	Phe	Asn	Phe	Ser	Leu	Lys	Leu	Leu	Thr	Cys	Leu	Leu	Tyr	Ile	
			100					105					110			
Val	Arg	Val	Leu	Leu	Asp	Asn	Pro	Asp	Gln	Gly	Ile	Gly	Cys	Trp	Gly	
		115					120					125				
Cys	Thr	Lys	Tyr	Asn	Tyr	Thr	Phe	Asn	Gly	Ser	Ser	Ser	Glu	Phe	His	
	130					135					140					
Trp	Ala	Pro	Ile	Leu	Trp	Val	Glu	Arg	Lys	Met	Ala	Leu	Trp	Val	Ile	
	145				150				155						160	
Gln	Val	Ile	Val	Ala	Thr	Ile	Ser	Phe	Leu	Glu	Thr	Met	Leu	Leu	Ile	
				165					170						175	
Tyr	Leu	Ser	Tyr	Lys	Gly	Asn	Ile	Trp	Glu	Gln	Ile	Phe	His	Val	Ser	
			180					185					190			
Phe	Val	Leu	Glu	Met	Ile	Asn	Thr	Leu	Pro	Phe	Ile	Ile	Thr	Val	Phe	
	195						200					205				
Trp	Pro	Pro	Leu	Arg	Asn	Leu	Phe	Ile	Pro	Val	Phe	Leu	Asn	Cys	Trp	
	210					215					220					
Leu	Ala	Lys	His	Ala	Leu	Glu	Asn	Met	Ile	Asn	Asp	Phe	His	Arg	Ala	
	225				230				235						240	
Ile	Leu	Arg	Thr	Gln	Ser	Ala	Met	Phe	Asn	Gln	Val	Leu	Ile	Leu	Phe	
				245					250						255	
Cys	Thr	Leu	Leu	Cys	Leu	Val	Phe	Thr	Gly	Thr	Cys	Gly	Ile	Gln	His	
		260						265					270			
Leu	Glu	Arg	Ala	Gly	Gly	Asn	Leu	Asn	Leu	Leu	Thr	Ser	Phe	Tyr	Phe	
	275					280					285					
Cys	Ile	Val	Thr	Phe	Ser	Thr	Val	Gly	Phe	Gly	Asp	Val	Thr	Pro	Lys	
	290					295					300					
Ile	Trp	Pro	Ser	Gln	Leu	Leu	Val	Val	Ile	Leu	Ile	Cys	Val	Thr	Leu	
	305				310				315						320	
Val	Val	Leu	Pro	Leu	Gln	Phe	Glu	Glu	Leu	Val	Tyr	Leu	Trp	Met	Glu	
				325					330						335	
Arg	Gln	Lys	Ser	Gly	Gly	Asn	Tyr	Ser	Arg	His	Arg	Ala	Arg	Thr	Glu	
			340					345					350			
Lys	His	Val	Val	Leu	Cys	Val	Ser	Ser	Leu	Lys	Ile	Asp	Leu	Leu	Met	
		355					360					365				
Asp	Phe	Leu	Asn	Glu	Phe	Tyr	Ala	His	Pro	Arg	Leu	Gln	Asp	Tyr	Tyr	
	370					375					380					
Val	Val	Ile	Leu	Cys	Pro	Ser	Glu	Met	Asp	Val	Gln	Val	Arg	Arg	Val	
	385				390				395						400	
Leu	Gln	Ile	Pro	Leu	Trp	Ser	Gln	Arg	Val	Ile	Tyr	Leu	Gln	Gly	Ser	
				405					410					415		
Ala	Leu	Lys	Asp	Gln	Asp	Leu	Met	Arg	Ala	Lys	Met	Asp	Asn	Gly	Glu	
			420					425					430			
Ala	Cys	Phe	Ile	Leu	Ser	Ser	Arg	Asn	Glu	Val	Asp	Arg	Thr	Ala	Ala	
		435					440					445				
Asp	His	Gln	Thr	Ile	Leu	Arg	Ala	Trp	Ala	Val	Lys	Asp	Phe	Ala	Pro	
	450					455						460				
Asn	Cys	Pro	Leu	Tyr	Val	Gln	Ile	Leu	Lys	Pro	Glu	Asn	Lys	Phe	His	
	465				470				475						480	
Val	Lys	Phe	Ala	Asp	His	Val	Val	Cys	Glu	Glu	Cys	Lys	Tyr	Ala		
				485					490					495		
Met	Leu	Ala	Leu	Asn	Cys	Ile	Cys	Pro	Ala	Thr	Ser	Thr	Leu	Ile	Thr	
			500					505					510			
Leu	Leu	Val	His	Thr	Ser	Arg	Gly	Gln	Glu	Gly	Gln	Glu	Ser	Pro	Glu	
		515					520					525				
Gln	Trp	Gln	Arg	Met	Tyr	Gly	Arg	Cys	Ser	Gly	Asn	Glu	Val	Tyr	His	
	530					535					540					
Ile	Arg	Met	Gly	Asp	Ser	Lys	Phe	Phe	Arg	Glu	Tyr	Glu	Gly	Lys	Ser	
	545				550					555					560	

Phe	Thr	Tyr	Ala	Ala	Phe	His	Ala	His	Lys	Lys	Tyr	Gly	Val	Cys	Leu	
				565					570					575		
Ile	Gly	Leu	Lys	Arg	Glu	Glu	Asn	Lys	Ser	Ile	Leu	Leu	Asn	Pro	Gly	
			580					585					590			
Pro	Arg	His	Ile	Leu	Ala	Ala	Ser	Asp	Thr	Cys	Phe	Tyr	Ile	Asn	Ile	
		595					600					605				
Thr	Lys	Glu	Glu	Asn	Ser	Ala	Phe	Ile	Phe	Lys	Gln	Glu	Glu	Lys	Gln	
	610					615					620					
Asn	Arg	Arg	Gly	Leu	Ala	Gly	Gln	Ala	Leu	Tyr	Glu	Gly	Pro	Ser	Arg	
625					630					635					640	
Leu	Pro	Val	His	Ser	Ile	Ile	Ala	Ser	Met	Val	Ala	Met	Asp	Leu	Gln	
			645						650					655		
Asn	Thr	Asp	Cys	Arg	Pro	Ser	Gln	Gly	Gly	Ser	Gly	Gly	Gly	Gly	Gly	
		660						665					670			
Lys	Leu	Thr	Leu	Pro	Thr	Glu	Asn	Gly	Ser	Gly	Ser	Arg	Arg	Pro	Ser	
		675				680						685				
Ile	Ala	Pro	Val	Leu	Glu	Leu	Ala	Asp	Ser	Ser	Ala	Leu	Leu	Pro	Cys	
	690					695					700					
Asp	Leu	Leu	Ser	Asp	Gln	Ser	Glu	Asp	Glu	Val	Thr	Pro	Ser	Asp	Asp	
705					710					715					720	
Glu	Gly	Leu	Ser	Val	Val	Glu	Tyr	Val	Lys	Gly	Tyr	Pro	Pro	Asn	Ser	
			725						730					735		
Pro	Tyr	Ile	Gly	Ser	Ser	Pro	Thr	Leu	Cys	His	Leu	Leu	Pro	Val	Lys	
			740					745						750		
Ala	Pro	Phe	Cys	Cys	Leu	Arg	Leu	Asp	Lys	Gly	Cys	Lys	His	Asn	Ser	
		755					760					765				
Tyr	Glu	Asp	Ala	Lys	Ala	Tyr	Gly	Phe	Lys	Asn	Lys	Leu	Ile	Ile	Val	
	770					775					780					
Ser	Ala	Glu	Thr	Ala	Gly	Asn	Gly	Leu	Tyr	Asn	Phe	Ile	Val	Pro	Leu	
785					790					795					800	
Arg	Ala	Tyr	Tyr	Arg	Ser	Arg	Arg	Glu	Leu	Asn	Pro	Ile	Val	Leu	Leu	
				805					810					815		
Leu	Asp	Asn	Lys	Pro	Asp	His	His	Phe	Leu	Glu	Ala	Ile	Cys	Cys	Phe	
		820						825					830			
Pro	Met	Val	Tyr	Tyr	Met	Glu	Gly	Ser	Val	Asp	Asn	Leu	Asp	Ser	Leu	
		835					840					845				
Leu	Gln	Cys	Gly	Ile	Ile	Tyr	Ala	Asp	Asn	Leu	Val	Val	Val	Asp	Lys	
	850					855					860					
Glu	Ser	Thr	Met	Ser	Ala	Glu	Glu	Asp	Tyr	Met	Ala	Asp	Ala	Lys	Thr	
865					870					875					880	
Ile	Val	Asn	Val	Gln	Thr	Met	Phe	Arg	Leu	Phe	Pro	Ser	Leu	Ser	Ile	
				885					890					895		
Thr	Thr	Glu	Leu	Thr	His	Pro	Ser	Asn	Met	Arg	Phe	Met	Gln	Phe	Arg	
		900						905					910			
Ala	Lys	Asp	Ser	Tyr	Ser	Leu	Ala	Leu	Ser	Lys	Leu	Glu	Lys	Gln	Glu	
		915					920						925			
Arg	Glu	Asn	Gly	Ser	Asn	Leu	Ala	Phe	Met	Phe	Arg	Leu	Pro	Phe	Ala	
	930					935					940					
Ala	Gly	Arg	Val	Phe	Ser	Ile	Ser	Met	Leu	Asp	Thr	Leu	Leu	Tyr	Gln	
945					950					955					960	
Ser	Phe	Val	Lys	Asp	Tyr	Met	Ile	Thr	Ile	Thr	Arg	Leu	Leu	Leu	Gly	
			965						970					975		
Leu	Asp	Thr	Thr	Pro	Gly	Ser	Gly	Tyr	Leu	Cys	Ala	Met	Lys	Val	Thr	
		980						985					990			
Glu	Asp	Asp	Leu	Trp	Ile	Arg	Thr	Tyr	Gly	Arg	Leu	Phe	Gln	Lys	Leu	
		995					1000						1005			
Cys	Ser	Ser	Ser	Ala	Glu	Ile	Pro	Ile	Gly	Ile	Tyr	Arg	Thr	Glu	Cys	
	1010					1015						1020				
His	Val	Phe	Ser	Ser	Glu	Pro	His	Asp	Leu	Arg	Ala	Gln	Ser	Gln	Ile	
1025					1030					1035					1040	

Ser Val Asn Met Glu Asp Cys Glu Asp Thr Arg Glu Ala Lys Gly Pro
 1045 1050 1055
 Trp Gly Thr Arg Ala Ala Ser Gly Gly Gly Ser Thr His Gly Arg His
 1060 1065 1070
 Gly Gly Ser Ala Asp Pro Val Glu His Pro Leu Leu Arg Arg Lys Ser
 1075 1080 1085
 Leu Gln Trp Ala Arg Lys Leu Ser Arg Lys Ser Ser Lys Gln Ala Gly
 1090 1095 1100
 Lys Ala Pro Met Thr Thr Asp Trp Ile Thr Gln Arg Leu Ser Leu
 1105 1110 1115 1120
 Tyr Arg Arg Ser Glu Arg Gln Glu Leu Ser Glu Leu Val Lys Asn Arg
 1125 1130 1135
 Met Lys His Leu Gly Leu Pro Thr Thr Gly Tyr Glu Asp Val Ala Asn
 1140 1145 1150
 Leu Thr Ala Ser Asp Val Met Asn Arg Val Asn Leu Gly Tyr Leu Gln
 1155 1160 1165
 Asp Glu Met Asn Asp His His Gln Asn Thr Leu Ser Tyr Val Leu Ile
 1170 1175 1180
 Asn Pro Pro Pro Asp Thr Arg Leu Glu Pro Asn Asp Ile Val Tyr Leu
 1185 1190 1195 1200
 Ile Arg Ser Asp Pro Leu Ala His Val Thr Ser Ser Ser Gln Ser Arg
 1205 1210 1215
 Lys Ser Ser Cys Ser Asn Lys Leu Ser Ser Cys Asn Pro Glu Thr Arg
 1220 1225 1230
 Asp Glu Thr Gln Leu
 1235

<210> 33

<211> 638

<212> PRT

<213> Rattus norvegicus

<400> 33

Met Gly Lys Ile Glu Asn Asn Glu Arg Val Ile Leu Asn Val Gly Gly
 1 5 10 15
 Thr Arg His Glu Thr Tyr Arg Ser Thr Leu Lys Thr Leu Pro Gly Thr
 20 25 30
 Arg Leu Ala Leu Leu Ala Ser Ser Glu Pro Gln Gly Asp Cys Leu Thr
 35 40 45
 Ala Ala Gly Asp Lys Leu Gln Pro Leu Pro Pro Pro Leu Ser Pro Pro
 50 55 60
 Pro Arg Pro Pro Pro Leu Ser Pro Val Pro Ser Gly Cys Phe Glu Gly
 65 70 75 80
 Gly Ala Gly Asn Cys Ser Ser His Gly Gly Asn Gly Ser Asp His Pro
 85 90 95
 Gly Gly Gly Arg Glu Phe Phe Phe Asp Arg His Pro Gly Val Phe Ala
 100 105 110
 Tyr Val Leu Asn Tyr Tyr Arg Thr Gly Lys Leu His Cys Pro Ala Asp
 115 120 125
 Val Cys Gly Pro Leu Phe Glu Glu Leu Ala Phe Trp Gly Ile Asp
 130 135 140
 Glu Thr Asp Val Glu Pro Cys Cys Trp Met Thr Tyr Arg Gln His Arg
 145 150 155 160
 Asp Ala Glu Glu Ala Leu Asp Ile Phe Glu Thr Pro Asp Leu Ile Gly
 165 170 175
 Gly Asp Pro Gly Asp Asp Glu Asp Leu Gly Gly Lys Arg Leu Gly Ile
 180 185 190
 Glu Asp Ala Ala Gly Leu Gly Gly Pro Asp Gly Lys Ser Gly Arg Trp
 195 200 205

Arg	Lys	Leu	Gln	Pro	Arg	Met	Trp	Ala	Leu	Phe	Glu	Asp	Pro	Tyr	Ser
210						215					220				
Ser	Arg	Ala	Ala	Arg	Phe	Ile	Ala	Phe	Ala	Ser	Leu	Phe	Phe	Ile	Leu
225					230					235					240
Val	Ser	Ile	Thr	Thr	Phe	Cys	Leu	Glu	Thr	His	Glu	Ala	Phe	Asn	Ile
				245					250					255	
Val	Lys	Asn	Lys	Thr	Glu	Pro	Val	Ile	Asn	Gly	Thr	Ser	Ala	Val	Leu
			260					265					270		
Gln	Tyr	Glu	Ile	Glu	Thr	Asp	Pro	Ala	Leu	Thr	Tyr	Val	Glu	Gly	Val
		275					280					285			
Cys	Val	Val	Trp	Phe	Thr	Phe	Glu	Phe	Leu	Val	Arg	Ile	Val	Phe	Ser
	290					295					300				
Pro	Asn	Lys	Leu	Glu	Phe	Ile	Lys	Asn	Leu	Leu	Asn	Ile	Ile	Asp	Phe
305					310					315					320
Val	Ala	Ile	Leu	Pro	Phe	Tyr	Leu	Glu	Val	Gly	Leu	Ser	Gly	Leu	Ser
				325					330					335	
Ser	Lys	Ala	Ala	Lys	Asp	Val	Leu	Gly	Phe	Leu	Arg	Val	Val	Arg	Phe
			340					345					350		
Val	Arg	Ile	Leu	Arg	Ile	Phe	Lys	Leu	Thr	Arg	His	Phe	Val	Gly	Leu
		355					360					365			
Arg	Val	Leu	Gly	His	Thr	Leu	Arg	Ala	Ser	Thr	Asn	Glu	Phe	Leu	Leu
	370					375					380				
Leu	Ile	Ile	Phe	Leu	Ala	Leu	Gly	Val	Leu	Ile	Phe	Ala	Thr	Met	Ile
385					390					395					400
Tyr	Tyr	Ala	Glu	Arg	Val	Gly	Ala	Gln	Pro	Asn	Asp	Pro	Ser	Ala	Ser
				405					410					415	
Glu	His	Thr	Gln	Phe	Lys	Asn	Ile	Pro	Ile	Gly	Phe	Trp	Trp	Ala	Val
			420					425					430		
Val	Thr	Met	Thr	Thr	Leu	Gly	Tyr	Gly	Asp	Met	Tyr	Pro	Gln	Thr	Trp
		435					440					445			
Ser	Gly	Met	Leu	Val	Gly	Ala	Leu	Cys	Ala	Leu	Ala	Gly	Val	Leu	Thr
	450					455					460				
Ile	Ala	Met	Pro	Val	Pro	Val	Ile	Val	Asn	Asn	Phe	Gly	Met	Tyr	Tyr
465					470					475					480
Ser	Leu	Ala	Met	Ala	Lys	Gln	Lys	Leu	Pro	Arg	Lys	Arg	Lys	Lys	His
				485					490					495	
Ile	Pro	Pro	Ala	Pro	Leu	Ala	Ser	Ser	Pro	Thr	Phe	Cys	Lys	Thr	Glu
			500					505					510		
Leu	Asn	Met	Ala	Cys	Asn	Ser	Thr	Gln	Ser	Asp	Thr	Cys	Leu	Gly	Lys
		515					520					525			
Glu	Asn	Arg	Leu	Leu	Glu	His	Asn	Arg	Ser	Val	Leu	Ser	Gly	Asp	Asp
	530					535					540				
Ser	Thr	Gly	Ser	Glu	Pro	Pro	Leu	Ser	Pro	Pro	Glu	Arg	Leu	Pro	Ile
545					550					555					560
Arg	Arg	Ser	Ser	Thr	Arg	Asp	Lys	Asn	Arg	Arg	Gly	Glu	Thr	Cys	Phe
				565					570					575	
Leu	Leu	Thr	Thr	Gly	Asp	Tyr	Thr	Cys	Ala	Ser	Asp	Gly	Gly	Ile	Arg
			580					585					590		
Lys	Gly	Tyr	Glu	Lys	Ser	Arg	Ser	Leu	Asn	Asn	Ile	Ala	Gly	Leu	Ala
		595					600					605			
Gly	Asn	Ala	Leu	Arg	Leu	Ser	Pro	Val	Thr	Ser	Pro	Tyr	Asn	Ser	Pro
	610					615					620				
Cys	Pro	Leu	Arg	Arg	Ser	Arg	Ser	Pro	Ile	Pro	Ser	Ile	Leu		
625					630						635				

<210> 34

<211> 1187

<212> PRT

<213> Mus musculus

<400> 34

Met	Asp	Cys	Ser	Leu	Leu	Arg	Thr	Leu	Val	Arg	Arg	Tyr	Cys	Ala	Gly
1				5					10					15	
Glu	Glu	Asn	Trp	Val	Asp	Ser	Arg	Thr	Ile	Tyr	Val	Gly	His	Lys	Glu
			20					25					30		
Pro	Pro	Pro	Gly	Ala	Glu	Ala	Tyr	Ile	Pro	Gln	Arg	Tyr	Pro	Asp	Asn
		35					40					45			
Arg	Ile	Val	Ser	Ser	Lys	Tyr	Thr	Phe	Trp	Asn	Phe	Ile	Pro	Lys	Asn
		50				55					60				
Leu	Phe	Glu	Gln	Phe	Arg	Arg	Ile	Ala	Asn	Phe	Tyr	Phe	Leu	Ile	Ile
65					70					75					80
Phe	Leu	Val	Gln	Leu	Ile	Ile	Asp	Thr	Pro	Thr	Ser	Pro	Val	Thr	Ser
			85						90					95	
Gly	Leu	Pro	Leu	Phe	Phe	Val	Ile	Thr	Val	Thr	Ala	Ile	Lys	Gln	Gly
			100					105					110		
Tyr	Glu	Asp	Trp	Leu	Arg	His	Lys	Ala	Asp	Asn	Ala	Met	Asn	Gln	Cys
		115					120					125			
Pro	Val	His	Phe	Ile	Gln	His	Gly	Lys	Leu	Val	Arg	Lys	Gln	Ser	Arg
		130					135				140				
Lys	Leu	Arg	Val	Gly	Asp	Ile	Val	Met	Val	Lys	Glu	Asp	Glu	Thr	Phe
145					150					155					160
Pro	Cys	Asp	Leu	Ile	Phe	Leu	Ser	Ser	Asn	Arg	Ala	Asp	Gly	Thr	Cys
			165						170					175	
His	Val	Thr	Thr	Ala	Ser	Leu	Asp	Gly	Glu	Ser	Ser	His	Lys	Thr	His
			180					185					190		
Tyr	Ala	Val	Gln	Asp	Thr	Lys	Gly	Phe	His	Thr	Glu	Ala	Asp	Val	Asp
		195					200					205			
Ser	Leu	His	Ala	Thr	Ile	Glu	Cys	Glu	Gln	Pro	Gln	Pro	Asp	Leu	Tyr
		210					215				220				
Lys	Phe	Val	Gly	Arg	Ile	Asn	Val	Tyr	Asn	Asp	Leu	Asn	Asp	Pro	Val
225					230				235						240
Val	Arg	Pro	Leu	Gly	Ser	Glu	Asn	Leu	Leu	Leu	Arg	Gly	Ala	Thr	Leu
			245						250					255	
Lys	Asn	Thr	Glu	Lys	Ile	Phe	Gly	Val	Ala	Ile	Tyr	Thr	Gly	Met	Glu
			260					265					270		
Thr	Lys	Met	Ala	Leu	Asn	Tyr	Gln	Ser	Lys	Ser	Gln	Lys	Arg	Ser	Ala
		275					280					285			
Val	Glu	Lys	Ser	Met	Asn	Thr	Phe	Leu	Ile	Val	Tyr	Leu	Cys	Ile	Leu
		290					295				300				
Val	Ser	Lys	Ala	Leu	Ile	Asn	Thr	Val	Leu	Lys	Tyr	Val	Trp	Gln	Ser
305					310					315					320
Glu	Pro	Phe	Arg	Asp	Glu	Pro	Trp	Tyr	Asn	Glu	Lys	Thr	Glu	Ser	Glu
			325						330					335	
Arg	Gln	Arg	Asn	Leu	Phe	Leu	Arg	Ala	Phe	Thr	Asp	Phe	Leu	Ala	Phe
			340					345					350		
Met	Val	Leu	Phe	Asn	Tyr	Ile	Ile	Pro	Val	Ser	Met	Tyr	Val	Thr	Val
		355					360					365			
Glu	Met	Gln	Lys	Phe	Leu	Gly	Ser	Tyr	Phe	Ile	Thr	Trp	Asp	Glu	Asp
		370				375					380				
Met	Phe	Asp	Glu	Glu	Met	Gly	Glu	Gly	Pro	Leu	Val	Asn	Thr	Ser	Asp
385					390					395					400
Leu	Asn	Glu	Glu	Leu	Gly	Gln	Val	Glu	Tyr	Ile	Phe	Thr	Asp	Lys	Thr
			405						410					415	
Gly	Thr	Leu	Thr	Glu	Asn	Asn	Met	Ala	Phe	Lys	Glu	Cys	Cys	Ile	Glu
			420					425					430		
Gly	His	Val	Tyr	Val	Pro	His	Val	Ile	Cys	Asn	Gly	Gln	Val	Leu	Pro
		435					440					445			
Asp	Ser	Ser	Gly	Ile	Asp	Met	Ile	Asp	Ser	Ser	Pro	Gly	Val	Cys	Gly
		450				455					460				
Arg	Glu	Arg	Glu	Glu	Leu	Phe	Phe	Arg	Ala	Ile	Cys	Leu	Cys	His	Thr
465					470					475					480

Val	Gln	Val	Lys	Asp	Asp	His	Cys	Gly	Asp	Asp	Val	Asp	Gly	Pro	Gln
				485					490					495	
Lys	Ser	Pro	Asp	Ala	Lys	Ser	Cys	Val	Tyr	Ile	Ser	Ser	Ser	Pro	Asp
			500					505					510		
Glu	Val	Ala	Leu	Val	Glu	Gly	Val	Gln	Arg	Leu	Gly	Phe	Thr	Tyr	Leu
		515					520					525			
Arg	Leu	Lys	Asp	Asn	Tyr	Met	Glu	Ile	Leu	Asn	Arg	Glu	Asn	Asp	Ile
	530					535					540				
Glu	Arg	Phe	Glu	Leu	Leu	Glu	Val	Leu	Thr	Phe	Asp	Ser	Val	Arg	Arg
	545					550				555					560
Arg	Met	Ser	Val	Ile	Val	Lys	Ser	Thr	Thr	Gly	Glu	Ile	Tyr	Leu	Phe
				565					570					575	
Cys	Lys	Gly	Ala	Asp	Ser	Ser	Ile	Phe	Pro	Arg	Val	Ile	Glu	Gly	Lys
			580					585					590		
Val	Asp	Gln	Val	Arg	Ser	Arg	Val	Glu	Arg	Asn	Ala	Val	Glu	Gly	Leu
	595						600					605			
Arg	Thr	Leu	Cys	Val	Ala	Tyr	Lys	Arg	Leu	Glu	Pro	Glu	Gln	Tyr	Glu
	610					615					620				
Asp	Ala	Cys	Arg	Leu	Leu	Gln	Ser	Ala	Lys	Val	Ala	Leu	Gln	Asp	Arg
	625				630					635					640
Glu	Lys	Lys	Leu	Ala	Glu	Ala	Tyr	Glu	Gln	Ile	Glu	Lys	Asp	Leu	Val
				645					650					655	
Leu	Leu	Gly	Ala	Thr	Ala	Val	Glu	Asp	Arg	Leu	Gln	Glu	Lys	Ala	Ala
			660					665					670		
Asp	Thr	Ile	Glu	Ala	Leu	Gln	Lys	Ala	Gly	Ile	Lys	Val	Trp	Val	Leu
	675						680					685			
Thr	Gly	Asp	Lys	Met	Glu	Thr	Ala	Ser	Ala	Thr	Cys	Tyr	Ala	Cys	Lys
	690				695					700					
Leu	Phe	Arg	Arg	Ser	Thr	Gln	Leu	Leu	Glu	Leu	Thr	Thr	Lys	Lys	Leu
	705				710					715					720
Glu	Glu	Gln	Ser	Leu	His	Asp	Val	Leu	Phe	Asp	Leu	Ser	Lys	Thr	Val
				725					730					735	
Leu	Arg	Cys	Ser	Gly	Ser	Met	Thr	Arg	Asp	Ser	Phe	Ser	Gly	Leu	Ser
			740					745					750		
Thr	Asp	Met	His	Asp	Tyr	Gly	Leu	Ile	Ile	Asp	Gly	Ala	Ala	Leu	Ser
	755						760					765			
Leu	Ile	Met	Lys	Pro	Arg	Glu	Asp	Gly	Ser	Ser	Ser	Gly	Asn	Tyr	Arg
	770					775					780				
Glu	Leu	Phe	Leu	Glu	Ile	Cys	Arg	Asn	Cys	Ser	Ala	Val	Leu	Cys	Cys
	785				790					795					800
Arg	Met	Ala	Pro	Leu	Gln	Lys	Ala	Gln	Ile	Val	Lys	Leu	Ile	Lys	Phe
				805					810					815	
Ser	Lys	Glu	His	Pro	Ile	Thr	Leu	Ala	Ile	Gly	Asp	Gly	Ala	Asn	Asp
			820					825					830		
Val	Ser	Met	Ile	Leu	Glu	Ala	His	Val	Gly	Ile	Gly	Val	Ile	Gly	Lys
		835					840					845			
Glu	Gly	Arg	Gln	Ala	Ala	Arg	Asn	Ser	Asp	Tyr	Ala	Ile	Pro	Lys	Phe
	850					855					860				
Lys	His	Leu	Lys	Lys	Met	Leu	Leu	Val	His	Gly	His	Phe	Tyr	Tyr	Ile
	865				870					875					880
Arg	Ile	Ser	Glu	Leu	Val	Gln	Tyr	Phe	Phe	Tyr	Lys	Asn	Val	Cys	Phe
				885					890					895	
Ile	Phe	Pro	Gln	Phe	Leu	Tyr	Gln	Phe	Phe	Cys	Gly	Phe	Ser	Gln	Gln
			900					905					910		
Thr	Leu	Tyr	Asp	Thr	Ala	Tyr	Leu	Thr	Leu	Tyr	Asn	Ile	Ser	Phe	Thr
	915						920					925			
Ser	Leu	Pro	Ile	Leu	Leu	Tyr	Ser	Leu	Met	Glu	Gln	His	Val	Gly	Ile
	930					935					940				
Asp	Val	Leu	Lys	Arg	Asp	Pro	Thr	Leu	Tyr	Arg	Asp	Ile	Ala	Lys	Asn
	945				950					955					960

Ala Leu Leu Arg Trp Arg Val Phe Ile Tyr Trp Thr Phe Leu Gly Val
 965 970 975
 Phe Asp Ala Leu Val Phe Phe Phe Gly Ala Tyr Phe Ile Phe Glu Asn
 980 985 990
 Thr Thr Val Thr Ile Asn Gly Gln Met Phe Gly Asn Trp Thr Phe Gly
 995 1000 1005
 Thr Leu Val Phe Thr Val Met Val Leu Thr Val Thr Leu Lys Leu Ala
 1010 1015 1020
 Leu Asp Thr His Tyr Trp Thr Trp Ile Asn His Phe Val Ile Trp Gly
 1025 1030 1035 1040
 Ser Leu Leu Phe Tyr Ile Ala Phe Ser Leu Leu Trp Gly Gly Val Ile
 1045 1050 1055
 Trp Pro Phe Leu Ser Tyr Gln Arg Met Tyr Tyr Val Phe Ile Ser Met
 1060 1065 1070
 Leu Ser Ser Gly Pro Ala Trp Leu Gly Ile Ile Leu Leu Val Thr Val
 1075 1080 1085
 Gly Leu Leu Pro Asp Val Leu Lys Lys Val Leu Cys Arg Gln Leu Trp
 1090 1095 1100
 Pro Thr Ala Thr Glu Arg Thr Gln Asn Ile Gln His Gln Asp Ser Ile
 1105 1110 1115 1120
 Ser Glu Phe Thr Pro Leu Ala Ser Leu Pro Ser Trp Gly Ala Gln Gly
 1125 1130 1135
 Ser Arg Leu Leu Ala Ala Gln Cys Ser Ser Pro Ser Gly Arg Val Val
 1140 1145 1150
 Cys Ser Arg Trp Glu Ser Glu Glu Cys Pro Val Leu Pro Leu His Pro
 1155 1160 1165
 Gly Leu Pro His Lys Ala Arg Tyr Gly Cys Cys Arg Ser Ser Leu Glu
 1170 1175 1180
 Met Pro Thr
 1185

<210> 35

<211> 1508

<212> PRT

<213> Mus musculus

<400> 35

Met Glu Arg Glu Leu Pro Ala Ala Glu Glu Ser Ala Ser Ser Gly Trp
 1 5 10 15
 Arg Arg Pro Arg Arg Arg Arg Trp Glu Gly Arg Thr Arg Thr Val Arg
 20 25 30
 Ser Asn Leu Leu Pro Pro Leu Gly Thr Glu Asp Ser Thr Ile Gly Ala
 35 40 45
 Pro Lys Gly Glu Arg Leu Leu Met Arg Gly Cys Ile Gln His Leu Ala
 50 55 60
 Asp Asn Arg Leu Lys Thr Thr Lys Tyr Thr Leu Leu Ser Phe Leu Pro
 65 70 75 80
 Lys Asn Leu Phe Glu Gln Phe His Arg Leu Ala Asn Val Tyr Phe Val
 85 90 95
 Phe Ile Ala Leu Leu Asn Phe Val Pro Ala Val Asn Ala Phe Gln Pro
 100 105 110
 Gly Leu Ala Leu Ala Pro Val Leu Phe Ile Leu Ala Val Thr Ala Ile
 115 120 125
 Lys Asp Leu Trp Glu Asp Tyr Ser Arg His Arg Ser Asp His Glu Ile
 130 135 140
 Asn His Leu Gly Cys Leu Val Phe Ser Arg Glu Glu Lys Lys Tyr Val
 145 150 155 160
 Asn Arg Tyr Trp Lys Glu Ile Arg Val Gly Asp Phe Val Arg Leu Cys
 165 170 175

Cys	Asn	Glu	Ile	Ile	Pro	Ala	Asp	Ile	Leu	Leu	Leu	Ser	Ser	Ser	Asp		
			180					185					190				
Pro	Asp	Gly	Leu	Cys	His	Ile	Glu	Thr	Ala	Asn	Leu	Asp	Gly	Glu	Thr		
		195					200					205					
Asn	Leu	Lys	Arg	Arg	Gln	Val	Val	Arg	Gly	Phe	Ser	Glu	Leu	Val	Ser		
	210					215					220						
Glu	Phe	Asn	Pro	Leu	Thr	Phe	Thr	Ser	Val	Ile	Glu	Cys	Glu	Lys	Pro		
225					230					235					240		
Asn	Asn	Asp	Leu	Ser	Arg	Phe	Arg	Gly	Tyr	Ile	Met	His	Ser	Asn	Gly		
				245					250					255			
Glu	Lys	Ala	Gly	Leu	His	Lys	Glu	Asn	Leu	Leu	Leu	Arg	Gly	Cys	Thr		
			260					265					270				
Ile	Arg	Asn	Thr	Glu	Ala	Val	Ala	Gly	Ile	Val	Ile	Tyr	Ala	Gly	His		
		275					280					285					
Glu	Thr	Lys	Ala	Leu	Leu	Asn	Asn	Ser	Gly	Pro	Arg	Tyr	Lys	Arg	Ser		
	290					295					300						
Gln	Leu	Glu	Arg	Gln	Met	Asn	Cys	Asp	Val	Leu	Trp	Cys	Val	Leu	Leu		
305					310					315					320		
Leu	Val	Cys	Ile	Ser	Leu	Phe	Ser	Ala	Val	Gly	His	Gly	Leu	Trp	Val		
				325					330					335			
Arg	Arg	Tyr	Gln	Glu	Lys	Lys	Ala	Leu	Phe	Asp	Val	Pro	Glu	Ser	Asp		
			340					345					350				
Gly	Ser	Ser	Leu	Ser	Pro	Ala	Thr	Ala	Ala	Val	Tyr	Ser	Phe	Phe	Thr		
		355					360					365					
Met	Ile	Ile	Val	Leu	Gln	Val	Leu	Ile	Pro	Ile	Ser	Leu	Tyr	Val	Ser		
	370					375					380						
Ile	Glu	Ile	Val	Lys	Val	Cys	Gln	Val	Tyr	Phe	Ile	Asn	Gln	Asp	Ile		
385					390					395				400			
Glu	Leu	Tyr	Asp	Glu	Glu	Thr	Asp	Ser	Gln	Leu	Gln	Cys	Arg	Ala	Leu		
				405					410					415			
Asn	Ile	Thr	Glu	Asp	Leu	Gly	Gln	Ile	Lys	Tyr	Ile	Phe	Ser	Asp	Lys		
			420					425					430				
Thr	Gly	Thr	Leu	Thr	Glu	Asn	Lys	Met	Val	Phe	Arg	Arg	Cys	Thr	Val		
		435					440					445					
Ser	Gly	Ile	Glu	Tyr	Ser	His	Asp	Ala	Asn	Ala	Gln	Arg	Leu	Ala	Arg		
	450					455					460						
Tyr	Gln	Glu	Ala	Asp	Ser	Glu	Glu	Glu	Glu	Val	Val	Ser	Lys	Val	Gly		
465					470					475					480		
Thr	Ile	Ser	His	Arg	Gly	Ser	Thr	Gly	Ser	His	Gln	Ser	Ile	Trp	Met		
				485					490					495			
Thr	His	Lys	Thr	Gln	Ser	Ile	Lys	Ser	His	Arg	Arg	Thr	Gly	Ser	Arg		
			500					505					510				
Ala	Glu	Ala	Lys	Arg	Ala	Ser	Met	Leu	Ser	Lys	His	Thr	Ala	Phe	Ser		
		515					520					525					
Ser	Pro	Met	Glu	Lys	Asp	Ile	Thr	Pro	Asp	Pro	Lys	Leu	Leu	Glu	Lys		
	530					535					540						
Val	Ser	Glu	Cys	Asp	Arg	Phe	Leu	Ala	Ile	Ala	Arg	His	Gln	Glu	His		
545					550					555					560		
Pro	Leu	Ala	His	Leu	Ser	Pro	Glu	Leu	Ser	Asp	Val	Phe	Asp	Phe	Phe		
				565					570					575			
Ile	Ala	Leu	Thr	Ile	Cys	Asn	Thr	Val	Val	Val	Thr	Ser	Pro	Asp	Gln		
			580					585					590				
Pro	Arg	Gln	Lys	Val	Arg	Val	Arg	Phe	Glu	Leu	Lys	Ser	Pro	Val	Lys		
		595					600					605					
Thr	Ile	Glu	Asp	Phe	Leu	Arg	Arg	Phe	Thr	Pro	Ser	Arg	Leu	Ala	Ser		
	610					615					620						
Gly	Cys	Ser	Ser	Ile	Gly	Asn	Leu	Ser	Thr	Ser	Lys	Ser	Ser	His	Lys		
625					630					635					640		
Ser	Gly	Ser	Ala	Phe	Leu	Pro	Ser	Leu	Ser	Gln	Asp	Ser	Met	Leu	Leu		
				645					650					655			

Gly Leu Glu Glu Lys Leu Gly Gln Thr Ala Pro Ser Ile Ala Ser Asn
 660 665 670
 Gly Tyr Ala Ser Gln Ala Gly Gln Glu Glu Ser Trp Ala Ser Asp Cys
 675 680 685
 Thr Thr Asp Gln Lys Cys Pro Gly Glu Gln Arg Glu Gln Gln Glu Gly
 690 695 700
 Glu Leu Arg Tyr Glu Ala Glu Ser Pro Asp Glu Ala Ala Leu Val Tyr
 705 710 715 720
 Ala Ala Arg Ala Tyr Asn Cys Ala Leu Val Asp Arg Leu His Asp Gln
 725 730 735
 Val Ser Val Glu Leu Pro His Leu Gly Arg Leu Thr Phe Glu Leu Leu
 740 745 750
 His Thr Leu Gly Phe Asp Ser Ile Arg Lys Arg Met Ser Val Val Ile
 755 760 765
 Arg His Pro Leu Thr Asp Glu Ile Asn Val Tyr Thr Lys Gly Ala Asp
 770 775 780
 Ser Val Val Met Asp Leu Leu Leu Pro Cys Ser Ser Asp Asp Ala Arg
 785 790 795 800
 Gly Arg His Gln Lys Lys Ile Arg Ser Lys Thr Gln Asn Tyr Leu Asn
 805 810 815
 Leu Tyr Ala Val Glu Gly Leu Arg Thr Leu Cys Ile Ala Lys Arg Val
 820 825 830
 Leu Ser Lys Glu Glu Tyr Ala Cys Trp Leu Gln Ser His Ile Glu Ala
 835 840 845
 Glu Ala Ser Val Glu Ser Arg Glu Glu Leu Leu Phe Gln Ser Ala Val
 850 855 860
 Arg Leu Glu Thr Asn Leu His Leu Leu Gly Ala Thr Gly Ile Glu Asp
 865 870 875 880
 Arg Leu Gln Glu Gly Val Pro Glu Thr Ile Ala Lys Leu Arg Gln Ala
 885 890 895
 Gly Leu Gln Ile Trp Val Leu Thr Gly Asp Lys Gln Glu Thr Ala Ile
 900 905 910
 Asn Ile Ala Tyr Ala Cys Lys Leu Leu Asp His Gly Glu Glu Val Ile
 915 920 925
 Thr Leu Asn Ala Asp Ser Gln Glu Ala Cys Ala Ala Leu Leu Asp Gln
 930 935 940
 Cys Leu Ser Tyr Val Gln Ser Arg Asn Pro Arg Ser Thr Leu Gln Asn
 945 950 955 960
 Ser Glu Ser Asn Leu Ser Val Gly Phe Ser Phe Asn Pro Val Ser Thr
 965 970 975
 Ser Thr Asp Ala Ser Pro Ser Pro Ser Leu Val Ile Asp Gly Arg Ser
 980 985 990
 Leu Ala Tyr Ala Leu Glu Lys Ser Leu Glu Asp Lys Phe Leu Phe Leu
 995 1000 1005
 Ala Lys Gln Cys Arg Ser Val Leu Cys Cys Arg Ser Thr Pro Leu Gln
 1010 1015 1020
 Lys Ser Met Val Val Lys Leu Val Arg Ser Lys Leu Lys Ala Met Thr
 1025 1030 1035 1040
 Leu Ala Ile Gly Asp Gly Ala Asn Asp Val Ser Met Ile Gln Val Ala
 1045 1050 1055
 Asp Val Gly Val Gly Ile Ser Gly Gln Glu Gly Met Gln Ala Val Met
 1060 1065 1070
 Ala Ser Asp Phe Ala Val Pro Arg Phe Arg Tyr Leu Glu Arg Leu Leu
 1075 1080 1085
 Ile Val His Gly His Trp Cys Tyr Ser Arg Leu Ala Asn Met Val Leu
 1090 1095 1100
 Tyr Phe Phe Tyr Lys Asn Thr Met Ser Val Gly Leu Leu Phe Trp Phe
 1105 1110 1115 1120
 Gln Phe Tyr Cys Gly Phe Ser Ala Ser Ala Met Ile Asp Gln Trp Tyr
 1125 1130 1135

Leu Ile Phe Phe Asn Leu Leu Phe Ser Ser Leu Pro Gln Leu Val Thr
 1140 1145 1150
 Gly Val Leu Asp Lys Asp Val Pro Ala Asp Met Leu Leu Arg Glu Pro
 1155 1160 1165
 Gln Leu Tyr Lys Ser Gly Gln Asn Met Glu Glu Tyr Arg Pro Arg Ala
 1170 1175 1180
 Phe Trp Leu Asn Met Val Asp Ala Ala Phe Gln Ser Leu Val Cys Phe
 1185 1190 1195 1200
 Phe Ile Pro Tyr Leu Ala Tyr Tyr Asp Ser Asp Val Asp Val Phe Thr
 1205 1210 1215
 Trp Gly Thr Pro Val Thr Ala Ile Ala Leu Phe Thr Phe Leu Leu His
 1220 1225 1230
 Leu Gly Ile Glu Thr Lys Thr Trp Thr Trp Leu Asn Trp Leu Ala Cys
 1235 1240 1245
 Gly Phe Ser Thr Phe Leu Phe Phe Ser Val Ala Leu Ile Tyr Asn Thr
 1250 1255 1260
 Ser Cys Ala Thr Cys Tyr Pro Pro Ser Asn Pro Tyr Trp Thr Met Gln
 1265 1270 1275 1280
 Thr Leu Leu Gly Asp Pro Leu Phe Tyr Leu Thr Cys Leu Ile Ala Pro
 1285 1290 1295
 Ile Ala Ala Leu Leu Pro Arg Leu Phe Phe Lys Ala Leu Gln Gly Ser
 1300 1305 1310
 Leu Phe Pro Thr Gln Leu Gln Leu Gly Arg Gln Leu Ala Lys Lys Pro
 1315 1320 1325
 Leu Asn Lys Phe Ser Asp Pro Lys Glu Thr Phe Ala Gln Gly Gln Pro
 1330 1335 1340
 Pro Gly His Ser Glu Thr Glu Leu Ser Glu Arg Lys Thr Met Gly Pro
 1345 1350 1355 1360
 Phe Glu Thr Leu Pro Arg Asp Cys Ala Ser Gln Ala Ser Gln Phe Thr
 1365 1370 1375
 Gln Gln Leu Thr Cys Ser Pro Glu Ala Ser Gly Glu Pro Ser Ala Val
 1380 1385 1390
 Asp Thr Asn Met Pro Leu Arg Glu Asn Thr Leu Leu Glu Gly Leu Gly
 1395 1400 1405
 Ser Gln Ala Ser Gly Ser Ser Met Pro Arg Gly Ala Ile Ser Glu Val
 1410 1415 1420
 Cys Pro Gly Asp Ser Lys Arg Gln Ser Ser Ser Ala Ser Gln Thr Ala
 1425 1430 1435 1440
 Arg Leu Ser Ser Leu Phe His Leu Pro Ser Phe Gly Ser Leu Asn Trp
 1445 1450 1455
 Ile Ser Ser Leu Ser Leu Ala Ser Gly Leu Gly Ser Val Leu Gln Leu
 1460 1465 1470
 Ser Gly Ser Ser Leu Gln Met Asp Lys Gln Asp Gly Glu Phe Leu Ser
 1475 1480 1485
 Asn Pro Pro Gln Pro Glu Gln Asp Leu His Ser Phe Gln Gly Gln Val
 1490 1495 1500
 Thr Gly Tyr Leu
 1505

<210> 36

<211> 1095

<212> PRT

<213> Mus musculus

<220>

<221> VARIANT

<222> 801, 1005

<223> Xaa = any amino acid

<400> 36

Met Pro Leu Met Met Ser Glu Glu Gly Phe Glu Asn Asp Glu Ser Asp
 1 5 10 15
~~Tyr His Thr Leu Pro Arg Ala Arg Ile Thr Arg Arg Lys Arg Gly Leu~~
 20 25 30
 Glu Trp Phe Val Cys Gly Gly Trp Lys Phe Leu Cys Thr Ser Cys Cys
 35 40 45
 Asp Trp Leu Ile Asn Val Cys Gln Arg Lys Lys Glu Leu Lys Ala Arg
 50 55 60
 Thr Val Trp Leu Gly Cys Pro Glu Lys Cys Glu Glu Lys His Pro Arg
 65 70 75 80
 Asn Ser Ile Lys Asn Gln Lys Tyr Asn Val Phe Thr Phe Ile Pro Gly
 85 90 95
 Val Leu Tyr Glu Gln Phe Lys Phe Phe Leu Asn Leu Tyr Phe Leu Val
 100 105 110
 Val Ser Cys Ser Gln Phe Val Pro Ala Leu Lys Ile Gly Tyr Leu Tyr
 115 120 125
 Thr Tyr Trp Ala Pro Leu Gly Phe Val Leu Ala Val Thr Ile Ala Arg
 130 135 140
 Glu Ala Ile Asp Glu Phe Arg Arg Phe Gln Arg Asp Lys Glu Met Asn
 145 150 155 160
 Ser Gln Leu Tyr Ser Lys Leu Thr Val Arg Gly Lys Val Gln Val Lys
 165 170 175
 Ser Ser Asp Ile Gln Val Gly Asp Leu Ile Ile Val Glu Lys Asn Gln
 180 185 190
 Arg Ile Pro Ser Asp Met Val Phe Leu Arg Thr Ser Glu Lys Ala Gly
 195 200 205
 Ser Cys Phe Ile Arg Thr Asp Gln Leu Asp Gly Glu Thr Asp Trp Lys
 210 215 220
 Leu Lys Val Ala Val Ser Cys Thr Gln Arg Leu Pro Ala Leu Gly Asp
 225 230 235 240
 Leu Phe Ser Ile Ser Ala Tyr Val Tyr Ala Gln Lys Pro Gln Leu Asp
 245 250 255
 Ile His Ser Phe Glu Gly Thr Phe Thr Arg Glu Asp Ser Asp Pro Pro
 260 265 270
 Ile His Glu Ser Leu Ser Ile Glu Asn Thr Leu Trp Ala Ser Thr Ile
 275 280 285
 Val Ala Ser Gly Thr Val Ile Gly Val Val Ile Tyr Thr Gly Lys Glu
 290 295 300
 Thr Arg Ser Val Met Asn Thr Ser Asn Pro Asn Asn Lys Val Gly Leu
 305 310 315 320
 Leu Asp Leu Glu Leu Asn Gln Leu Thr Lys Ala Leu Phe Leu Ala Leu
 325 330 335
 Val Val Leu Ser Val Val Met Val Thr Leu Gln Gly Phe Ala Gly Pro
 340 345 350
 Trp Tyr Arg Asn Leu Phe Arg Phe Leu Leu Leu Phe Ser Tyr Ile Ile
 355 360 365
 Pro Ile Ser Leu Arg Val Asn Leu Asp Met Gly Lys Ala Ala Tyr Gly
 370 375 380
 Trp Met Ile Met Lys Asp Glu Asn Ile Pro Gly Thr Val Val Arg Thr
 385 390 395 400
 Ser Thr Ile Pro Glu Glu Leu Gly Arg Leu Val Tyr Leu Leu Thr Asp
 405 410 415
 Lys Thr Gly Thr Leu Thr Gln Asn Glu Met Val Phe Lys Arg Leu His
 420 425 430
 Leu Gly Thr Val Ser Tyr Gly Thr Asp Thr Met Asp Glu Ile Gln Ser
 435 440 445
 His Val Leu Asn Ser Tyr Leu Gln Val His Ser Gln Pro Ser Gly His
 450 455 460
 Asn Pro Ser Ser Ala Pro Leu Arg Arg Ser Gln Ser Ser Thr Pro Lys
 465 470 475 480

- 124 -


```

Phe Leu Ile Trp Val Leu Ile Ser Ile Tyr Gln Gly Gly Ile Leu Met
          965                      970                      975
Tyr Gly Ala Leu Leu Leu Phe Glu Asp Glu Phe Val His Val Val Ala
          980                      985                      990
Ile Ser Phe Thr Ala Leu Ile Leu Thr Glu Leu Leu Xaa Val Ala Leu
          995                      1000                      1005
Thr Ile Arg Thr Trp His Trp Leu Met Val Val Ala Glu Phe Leu Ser
          1010                      1015                      1020
Leu Gly Cys Tyr Val Ala Ser Leu Ala Phe Leu Asn Glu Tyr Phe Gly
          1025                      1030                      1035                      1040
Ile Gly Arg Val Ser Phe Gly Ala Phe Leu Asp Val Ala Phe Ile Thr
          1045                      1050                      1055
Thr Val Thr Phe Leu Trp Lys Val Ser Ala Ile Thr Val Val Ser Cys
          1060                      1065                      1070
Leu Pro Leu Tyr Val Leu Lys Tyr Leu Lys Arg Lys Leu Ser Pro Pro
          1075                      1080                      1085
Ser Tyr Ser Lys Leu Ser Ser
          1090                      1095

```

<210> 37
 <211> 9
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> domain

<220>
 <221> VARIANT
 <222> (1)
 <223> Xaa = Asp, Asn, or Ser

<220>
 <221> VARIANT
 <222> (2)
 <223> Xaa = Gln, Glu, Asn, or Arg

<220>
 <221> VARIANT
 <222> (3)
 <223> Xaa = Ser or Ala

<220>
 <221> VARIANT
 <222> (4)
 <223> Xaa = Leu, Ile, Val, Ser, Ala, or Asn

<220>
 <221> VARIANT
 <222> (5)
 <223> Xaa = Leu, Ile, or Val

<220>
 <221> VARIANT
 <222> (6)
 <223> Xaa = Thr, Ser, or Asn

<220>

<221> VARIANT

<222> (9)

<223> Xaa = Ser or Asn

<400> 37

Xaa Xaa Xaa Xaa Xaa Xaa Gly Glu Xaa

1

5

<210> 38

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> domain

<220>

<221> VARIANT

<222> (1)

<223> Xaa = Leu, Ile, or Val

<220>

<221> VARIANT

<222> (2)

<223> Xaa = Cys, Ala, Met, or Leu

<220>

<221> VARIANT

<222> (3)

<223> Xaa = Ser, Thr, Phe, or Leu

<220>

<221> VARIANT

<222> (9)

<223> Xaa = Leu or Ile

<400> 38

Xaa Xaa Xaa Asp Lys Thr Gly Thr Xaa Thr

1

5

10

<210> 39

<211> 11

<212> PRT

<213> Artificial Sequence

<220>

<223> domain

<220>

<221> VARIANT

<222> (1)

<223> Xaa = Thr, Ile, or Val

<220>

<221> VARIANT

<222> (5)

<223> Xaa = any amino acid

<220>

<221> VARIANT

<222> (8)

<223> Xaa = Ala, Ser, or Gly

<220>

<221> VARIANT

<222> (10)

<223> Xaa = Ala, Ser, or Val

<400> 39

Xaa Gly Asp Gly Xaa Asn Asp Xaa Pro Xaa Leu
1 5 10

<210> 40

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

<223> domain

<220>

<221> VARIANT

<222> (6)

<223> Xaa = Leu, Ile, Val, or Met

<220>

<221> VARIANT

<222> (7)

<223> Xaa = Thr or Ile

<400> 40

Asp Lys Thr Gly Thr Xaa Xaa
1 5